Supplementary Online Content†

Song Z, Baicker K. Effect of a workplace wellness program on employee health and economic outcomes: a randomized clinical trial. JAMA. doi:10.1001/jama.2019.3307
†

TABLE OF CONTENTS

Section 1: Program Background
eMethods 1. Description of Program Modules 2
eMethods 2. Initial Power Calculations 4
eMethods 3. Statistical Analysis 5

Section 2: Figures
eFigure 1. Location of Treatment and Control Worksites 12
eFigure 2. Treatment Assignment and Analytic Samples 13

Section 3: Tables
eTable 1. Timeline of the Workplace Wellness Intervention 14
eTable 2. Summary of Outcome Data Collected 15
eTable 3. Pre-Specified Outcomes 16
eTable 4. Characteristics of the Study Population 18
eTable 5. Average Participation Rates by Module 21
eTable 6. Mean Values and Effect of Program on Self-Reported Health 22
eTable 7. Mean Values and Effect of Program on Clinical Measures of Health 26
eTable 8. Mean Values and Effect of Program on Medical Spending and Utilization 28
eTable 9. Mean Values and Effect of Program on Prescription Drug Spending and Utilization 31
eTable 10. Mean Values and Effect of Program on Employment Outcomes 34
eTable 11. Heterogeneity Analyses of Key Pre-Specified Outcomes 35
eTable 12. Alternative Definitions of Participation 37
eTable 13. Exposure Weights Only 39
eTable 14. Logit Model Estimates for Binary Outcome Variables 40
eTable 15. Analysis of Potential Selection: Participants vs. Non-Participants in Program Modules 43
eTable 16. Analysis of Potential Selection: Participants vs. Non-Participants in Survey/Biometrics 44
eTable 17. Analysis of Potential Selection: Participants vs. Non-Participants (Observational Design) 45

† This appendix complements the analysis plan that has been publicly archived at the National Institutes of Health and American Economic Association clinical trials registries. We gratefully acknowledge Jose Zubizarreta for guidance and contributions to the statistical methodology behind sample weights. We thank David Molitor and Julian Reif for guidance on the statistical software for multiple inference adjustment created in the Illinois Workplace Wellness Study, which was similarly used in this study. We thank Ozlem Blakeley, Kathryn Clark, Jack Huang, Bethany Maylone, Harlan Pittell, and Yuanxiaoyue Yang for excellent research assistance and project management. We also thank our study partners, BJ's Wholesale Club and Wellness Workdays, for their collaboration and assistance.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods 1: Description of Program Modules

The workplace wellness program we evaluated was delivered over an 18-month period at 20 randomly-selected worksites of BJ's Wholesale Club, located along the Eastern United States. The program was delivered by an experienced workplace wellness vendor, Wellness Workdays, through eight program modules, each of which focused on a key area of health and wellness. The names, timeline, description, and incentives for participation included in each module are summarized as follows. Please note that in the terminology of the firm, "club" denotes a worksite. In this section, we maintain this terminology. In the remainder of this document and in the paper, we use the more general term "worksite" or "site."

Module 1: Take Charge of Your Health
Timeline: 2/23/2015-3/27/2015 (Round 1), 4/13/2015-5/15/2015 (Round 2)

Summary: These two five-week programs were presented as a series of webinars with corresponding PowerPoints designed to help employees who participate take their health care into their own hands. Topics covered included: how to choose a health plan and primary care physician; what to expect from a routine visit; routine tests and screenings and recommended frequencies; how to get the most from a doctor's visit; choosing generic medications over the corresponding brand name; staying healthy by eating well, staying active, sleeping enough, and managing stress; and when to use primary care vs urgent care vs the emergency room.

Incentive: Employees who completed the webinars and returned the verification form received a $25 BJ's gift card. Employees only received a gift card for completing Take Charge one time.

Module 2: Nutrition for a Lifetime
Timeline: 6/1/2015-7/10/2015

Summary: This six-week program was presented as a series of webinar presentations or in paper form designed to help employees who participate achieve and maintain a healthy weight for life through the four pillars of health: nutrition, exercise, stress management, and sleep. Topics covered included: the negatives consequences of chronic stress and poor sleep habits and techniques to manage stress and improve sleep; good nutrition, including an overview on the different food groups and the amounts of each recommended per day; reasons for making exercise a priority and how to get the most out of a workout; foods to limit and foods to increase in a diet; appropriate portion sizing, especially for weight loss and weight maintenance; and choosing the right fats and the importance of fiber.

Incentive: Employees who completed 5 out of 6 webinars and returned the verification form received a $50 BJ's gift card. Employees with Cigna coverage received an additional $150 Weight Management Reimbursement in the form of a BJ's gift card for completion.

Modules 3 and 4: Club Cardio Challenge Rounds 1 and 2
Timeline: 8/10/2015-9/25/2015 (Round 1), 9/26/2015-11/16/2015 (Round 2)

Summary: These two seven-week programs were exercise-based. Employees were asked to complete 20 minutes or more of cardiovascular exercise at least 3 days per week and log their activity.

Incentive: Employees who completed 6 of the 7 weeks in either round 1 or round 2 earned a $25 BJs gift card. Employees who completed 12 out of 14 weeks over both rounds were eligible to enter a raffle at their worksite for a Fitbit. Employees with Cigna coverage who completed 12 out of 14 weeks received a $150 fitness reimbursement from Cigna on top of the raffle entry and gift card. Worksites were also in competition with the top worksite based on % participation and the top worksite with the highest average weekly minutes of exercise reported each receiving a trophy to display in the worksite, winner buttons for employee lanyards, and bragging rights.

Module 5: Maintain Don't Gain
Timeline: 11/23/2015-12/21/2015
Summary: This four-week weight management program included tracking Team Member's weekly weight via a paper log or portal, weight management handouts, and the opportunity for weekly coaching with the Registered Dietitian. The program Tip sheet topics included: BMI Chart, Balancing Calories, Generic Eating Plan, Appropriate Exercise Recommendations, Healthy Holiday Recipes, Holiday Stress Tips, Eating Healthy at a Party, and Winter Exercise. The program helped employees track their weight each week and offered tips on how to add physical activity to a daily routine and substitutions for options with fewer calories when dining out.
Incentive: Employees who completed at least 3 out of the 4 weeks of weight tracking and returned the verification form received a $25 BJ's gift card.

Module 6: Power Down the Pressure
Timeline: 1/18/2016-2/19/2016
Summary: This four-week program encouraged employees to learn effective methods for managing stress by asking them to complete at least one activity from a list of options for the week for at least 3 days of the week. Week 1 was called "Unplug" and included activities such as refraining from watching TV for a day or having an electronic-free meal with family or friends. Week 2 was titled "Boost Your Mood" and included activities like doing a random act of kindness, getting 8 hours of sleep, or spending time with a friend. Week 3 was "Exercise" and asked employees to take a new exercise class or do a 30-minute workout/activity outdoors. The final week was called "Relaxation and Meditation" and encouraged employees to keep a stress journal, color, and meditate.
Incentive: Employees who completed all four weeks of the program by completing at least 3 days of stress management activities a week and returned the verification form received a $25 BJ's gift card.

Module 7: Weight Loss Boot Camp
Timeline: 3/14/2016-4/8/2016
Summary: This four-week program aimed to teach employees methods for losing weight. Each week, employees completed four activities (eating five or more servings of fruits and vegetables, exercising for at least 30 minutes, avoiding sweetened beverages, and weighing themselves weekly) a minimum number of days each week, from two days the first week up to five days the final week. Cigna and Wellness Workdays Health Coaches were available to participants. Employees with BJ Health Insurance were provided with a Cigna Health Coach. Employees with other insurance were provided with a Wellness Workdays Health Coach.
Incentive: Employees who completed all four weeks and returned the verification form received a $25 BJ's gift card.

Module 8: Movin' in May
Timeline: 5/1/2016-5/31/2016
Summary: This four-week program encouraged employees to exercise for at least 30 minutes 3 days per week and track their exercise. Suggested exercise list included: bike, walk, run, elliptical, cardio classes, weight lifting, play sports like basketball or soccer, take an exercise class, yoga, swimming, and more.
Incentive: Employees who completed all four weeks of the challenge and returned the verification form were entered to win one of two $250 Visa gift cards at their worksite.

eMethods 2. Initial Power Calculations

After randomization of the worksites (with the number of treatment sites limited by the study budget), we conducted initial power calculations before implementing this randomized clinical trial or collecting outcome data. We made the following power calculations based on the assumptions listed below.

The study randomized the workplace wellness program to 20 treatment worksites. The remaining 140 worksites served as controls and did not receive wellness programming. Twenty of these were primary control worksites in which we would go on to field the Personal Health Assessment survey and clinical biometrics screening. We collected administrative records on health insurance claims and employment outcomes from all treatment and control worksites. We estimated 100 employees per worksite. The table below shows several stylized examples of outcomes and the minimum detectable effect sizes assuming a power of 0.8 and an intracluster correlation (ρ) of 0.05. This ρ is a conservative estimate based on our calculations of age, hours worked, and wages in the employment data (ρ ranging from 0.01 to 0.05).

For these ex ante power calculations, data on outcome variables were derived from the National Health and Nutrition Examination Survey, the Behavioral Risk Factor Surveillance System, the Medical Expenditure Panel Survey, and commercial insurance claims, rather than the study population (although we restricted the samples in the surveys to match the demographics of our study sample to the extent possible). The estimates shown here are for the intention-to-treat analysis (effect of the availability of the wellness program). Assuming a take-up in the treatment worksites of 33%, the detectable effect size in the local average treatment effect analysis would be 3 times larger.

Outcome	Example measure	Study Data Source	Est. Mean	Std. Dev.	Detectable Effect Size	Estimate Source
Physical health	Systolic blood pressure	PHA, kiosk	122	15.3	3.4 mmHg	NHANES
Physical health	Body mass index	PHA, kiosk	29.0	6.8	1.5 kg/m2	NHANES
Self-reported health	Excellent/very good (%)	Survey	52.2	--	11.0 pct pts	BRFSS
Health care use	ED visits per year	Survey, claims	0.28	1.01	0.22 visits	ESI claims
Health care use	Preventive visits per year	Survey, claims	0.39	0.55	0.12 visits	ESI claims
Medical spending	Outpatient ($/mo.)	Claims	221	763	169 $/mo.	ESI claims
Worker productivity	Absentee days	BJ's records	4.3	11.7	2.6 days	MEPS

After obtaining data from this randomized clinical trial, our estimated means shown above were close to the actual observed means of these variables (see Tables 2-5).

eMethods 3. Statistical Analysis

This section presents further details of the pre-specified statistical analysis. This section is adapted from the Analysis Plan that was finalized and publicly archived prior to conducting the analysis. The Analysis Plan can be found on clinicaltrials.gov and the American Economic Association Randomized Clinical Trials Registry.[1]

A. Intent-to-treat analysis

In the intent-to-treat analysis at the individual level, our goal is to estimate the average effect of a worker being randomized into a treatment site vs. a control site on outcomes of interest. We use a model that includes a treatment indicator capturing whether an individual was employed at a treatment vs. a control worksite. Individual-level observations were weighted based on the share of the intervention period during which the individual was employed at BJ's – or exposure to the intervention (discussed below). The model aims to answer the question: what is the effect of offering an individual the opportunity to participate in a wellness program? It is worth noting that in our experimental setting, individuals who worked at a treatment worksite but did not elect to participate actively in any of the wellness programming may still be "exposed" to the intervention by, for example, seeing posters in the common areas, sampling the healthier food made available in break rooms, or hearing about activities from participants at the site.

$$Y_{ij} = \beta_0 + \beta_1 TREATMENT_j + \beta_2 \boldsymbol{X}_{ij} + \varepsilon_{ij} \quad (1)$$

In this representative estimating equation, Y_{ij} denotes an outcome of interest for individual *i* who is employed at worksite *j*, such as medical spending. $TREATMENT_j$ is a binary indicator of whether the individual's worksite was randomized into the treatment or control arm. A small share of employees (2.6%) appeared in more than one worksite during the study period. We defined each individual's treatment or control status using the status of the worksite where the individual was originally employed, given that subsequent movement between worksites could in theory be endogenous. Standard errors are clustered by the worksite.

The coefficient on $TREATMENT_j$ (β_1) indicates the effect of being randomized into a treatment worksite, or the intent-to-treat (ITT). The ITT estimate is informative for employers considering implementing a wellness program. $\boldsymbol{X}_{ij}$ represents a vector of covariates that may help improve precision as well as account for chance differences in characteristics between treatment and control groups. These include:

- Age indicators: younger than 20 years (omitted), 20-29, 30-39, 40-49, 50-59, and 60 or older.[2]
- Sex indicator: male (omitted), female
- Age-sex interactions

[1] Song Z, Baicker K. The Impact of Workplace Wellness on Health, Health Care, and Employment Outcomes: A Randomized Controlled Trial, Analysis Plan—Phase 1. 2018 Mar 15. (https://clinicaltrials.gov/ProvidedDocs/58/NCT03167658/Prot_SAP_000.pdf); Baicker K, Maylone B, Song Z. The Impact of Employee Wellness Programs: A Randomized Controlled Trial. American Economic Association RCT Registry. 2018 Mar 15. (https://www.socialscienceregistry.org/trials/586)

[2] These age categories differ slightly from those originally specified in the analysis plan in order to be consistent with the pre-specified heterogeneity analysis.

- Race: black (omitted), white, Hispanic, and other
- Employment characteristics (measured at the time of a worker's first appearance in the data): full-time vs. part-time, employee type (salaried vs. hourly), job category (sales vs. non-sales vs. other)

We used two sets of weights in our primary analysis. First, each individual was assigned an "exposure" weight based on the extent of his or her exposure to the wellness program (i.e. the treatment) during the study period. Many BJ's employees joined or left BJ's employment during the course of this 18-month intervention. Moreover, many worked far less than full time. Outcomes for individuals with minimal exposure to the intervention are unlikely to be responsive to their small amount of time spent in a treatment vs. control worksite. Exposure weights are one way to account for this; an alternative would be estimating a dose-response model. We calculated this exposure weight using data on duration of employment and hours worked provided by BJ's. We summed the number of hours actually worked during the treatment period and divided by the number of hours a full-time employee would have worked during the study period, with resulting weights between 0 and 1. For example, a half-time worker who was employed for half of the treatment period would be assigned a weight of 0.25. See the publicly archived Analysis Plan, Table 12A, for summary statistics on these weights in the control group. Due to the potential endogeneity with the treatment, we did not examine the distribution across the treatment group prior to conducting the analysis.

Second, given that the treatment and control groups were not perfectly balanced on the set of raw observable characteristics after randomization, we derived a second set of weights that achieve balance between treatment and control workers on age, sex, and race—attributes that are not plausibly affected by the intervention. These weights were constructed to balance the demographic characteristics between the treatment and control groups with minimum variance between the weights, and were calibrated to be representative of the demographic attributes of the entire study population. This method has been shown to perform better than a model-based approach that fits a propensity score.[3] See the publicly archived Analysis Plan, Table 12B, for summary statistics on the balance weights between the treatment and control groups. In primary analyses, we use a composite weight constructed by multiplying the exposure weights and the balance weights together. In secondary analyses, we reassess a set of key outcomes using only the exposure weights.

B. Local average treatment effect

While our ITT analysis explores the effect of being randomized into a treatment worksite, a related but distinct question is: what is the effect of participating in the wellness program on the outcomes of interest? This second question produces a different estimate because not all employees in treatment worksites chose to participate. Some may not have found the wellness program appealing. Because of this endogenous participation choice, comparing those who participate in treatment worksites to all employees in control worksites may produce biased estimates of the effect of participation. We therefore model the impact of participation on outcomes using an instrumental variables two-stage least squares (2SLS) specification:

[3] Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. Journal of the American Statistical Association. 2015 Sep;110(511):910-922; Wang X, Zubizarreta JR. Minimal approximately balancing weights: Asymptotic properties and practical considerations. Biometrika. 2017;103(1):1-22; Hirshberg DA, Zubizarreta JR. On two approaches to weighting in causal inference. Epidemiology. 2017;28(6):812-816.

$$Y_{ij} = \gamma_0 + \gamma_1 PARTICIPATION_{ij} + \gamma_2 \boldsymbol{X}_{ij} + \mu_{ij} \quad (2)$$

where the endogenous *PARTICIPATION* variable is estimated via the first stage regression:

$$PARTICIPATION_{ij} = \pi_0 + \pi_1 TREATMENT_j + \pi_2 \boldsymbol{X}_{ij} + v_{ij} \quad (3)$$

γ_1 is the local average treatment effect (LATE) of participating in the wellness program. The first stage table below shows the results of estimating equation (3) for alternative definitions of participation. Our preferred specification used the binary indicator for whether a person ever participated in any module of the wellness program during the study period. In alternative specifications, we applied alternative definitions of participation including an indicator for participating in 3 or more modules and a continuous measure of the number of modules completed.

If no one in the control group received the treatment, we might interpret the 2SLS LATE as a treatment on the treated (TOT). This is nearly true by definition because, by construction, the control worksites did not have access to the wellness program modules. However, because we assigned employees to worksites based on their location at the beginning of the study period or when they first appeared in the employment database, a few individuals who moved from control worksites to treatment worksites during the study period did receive an opportunity to participate in data collection and the program. This accounts for the fact that the control group means in the table below are nearly, but not exactly, zero.

First Stage Estimates

Table A. All Individuals

Definition of Participation	**All Individuals**		**Completed PHA Survey**		**Cigna Coverage**	
	Control Mean	Estimated First Stage	Control Mean	Estimated First Stage	Control Means	Estimated First Stage
	(1)	(2)	(3)	(4)	(5)	(6)
Completed any module (%) (primary definition)	0.4	57.00 (2.50) [0.00]	2.5	77.81 (2.00) [0.00]	0.5	63.51 (2.68) [0.00]
Number of modules completed (#)	0.0	2.39 (0.11) [0.00]	0.1	3.82 (0.17) [0.00]	0.0	2.97 (0.16) [0.00]
Completed three or more modules (%)	0.2	39.93 (1.99) [0.00]	2.0	63.19 (2.83) [0.00]	0.4	49.08 (2.91) [0.00]
Average incentive payments ($)	0.4	103.97 (5.54) [0.00]	3.2	168.90 (8.14) [0.00]	0.7	169.16 (10.28) [0.00]
N	28936	32973	1088	2168	6626	7631

Table B. Stably Employed Subsample

Definition of Participation	**Entire Subsample**		**Completed PHA Survey**		**Cigna Coverage**	
	Control Mean	Estimated First Stage	Control Mean	Estimated First Stage	Control Means	Estimated First Stage
	(1)	(2)	(3)	(4)	(5)	(6)
Completed any module (%) (primary definition)	0.4	61.73 (2.63) [0.00]	2.7	81.35 (2.08) [0.00]	0.5	64.36 (2.84) [0.00]
Number of modules completed (#)	0.0	2.76 (0.13) [0.00]	0.1	4.23 (0.19) [0.00]	0.0	3.08 (0.16) [0.00]
Completed three or more modules (%)	0.3	45.90 (2.30) [0.00]	2.1	69.42 (2.87) [0.00]	0.4	50.51 (2.89) [0.00]
Average incentive payments ($)	0.5	126.02 (6.91) [0.00]	3.3	195.41 (9.76) [0.00]	0.7	187.73 (10.64) [0.00]
N	13442	15344	676	1353	5222	6016

Notes: Table A shows first stage estimates for all individuals. Table B shows first stage estimates for the stably employed subsample. In both tables, odd-numbered columns report control means for different definitions of *PARTICIPATION* and even-numbered columns report the estimated first stage coefficients (with standard errors in parentheses and p-values in brackets) on TREATMENT using the specified definition of PARTICIPATION (equation 3). Given that participation is the only channel through which incentive payments can be earned, we also include average incentive payments in these tables, though we do not instrument for it. All regressions include demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, full-time status, paid hourly status, and job category) and cluster standard errors by worksite. All samples other than the sample with Cigna coverage also control for whether or not the employee ever had Cigna coverage during the study period. Employee-level regressions are weighted by individual exposure to the treatment (exposure weights). In Table A, among the subset of individuals who completed a personal health assessment (PHA) survey (columns 3 and 4), there were 68 individuals initially assigned to control worksites who completed a minimum of 1 program module, thus giving us the approximately 2% of control means. The majority of these individuals had moved into a treatment worksite during the study period. In Table B, 58 of those individuals were in columns 3 and 4.

C. Addressing the inclusion of multiple related outcomes

We have multiple measures that capture closely related outcomes. This introduces two issues: first, combining information from these metrics may increase power. Second, we need to account for the multiple estimates of closely related outcomes in our inferential statistics.

We assessed groups of related outcomes by pre-specifying three standardized treatment effects. Specifically, we generated standardized treatment effects for each of the following groups:

- Biometrics (systolic and diastolic BP, cholesterol, HDL, glucose, BMI)
- Health behaviors (all PHA outcomes except emotional health and medical utilization)
- Mental health and well-being (all of the mental health and well-being outcomes in the PHA)

We conduct multiple inference adjustment within domains of outcomes. We adjusted for the number of outcomes tested within domains, largely as defined by the outcomes grouped within a particular table.

For each outcome, we report standard, per-comparison p-values and adjusted "family-wise" p-values that take into account the multiple related outcomes we pre-specified within each outcome category. The adjusted p-value speaks to the probability of rejecting the null hypothesis (i.e. no effect of the intervention on the category overall) on a given outcome under the null hypothesis that the intervention had no effect on any of the outcomes in that category. We used the Westfall and Young method for calculating these adjusted p-values (which, unlike the Bonferroni method, does not assume independence across the outcomes within a category).[4]

D. Pre-specified subgroup analyses

We performed two subgroup analyses at the individual level. We assessed differences in the effect of the wellness program by age and sex—two dimensions along which we observed fairly large differences in means in the control group at baseline in the control group, as shown below.

	Absenteeism (%)		**Total medical spending ($)**		**Regular exercise (%)**		**Considering losing weight (%)**	
	N	Control Mean	N	Control Mean	N	Control Mean	N	Control Mean
Gender								
Female	13339	2.86	3005	4740.30	517	56.88	526	61.53
Male	15597	2.44	3621	3354.24	394	68.07	371	49.21
Age								
Below 40	18726	2.48	2691	2397.55	504	67.59	480	54.88
40 and above	10210	2.80	3935	5009.16	407	56.47	417	57.46

	SF-8 physical summary score		**SF-8 mental summary score**		**Sweetened Drinks (No.)**		**BMI**		**Systolic BP (mmHg)**	
	N	Control Mean	N	Control Mean	N	Control Mean	N	Control Mean	N	Control Mean
Gender										
Female	539	49.93	539	50.16	566	1.66	609	30.31	612	120.59
Male	408	51.87	408	52.46	427	2.07	461	28.94	462	128.99
Age										
Below 40	515	51.97	515	49.65	537	2.14	578	29.13	581	119.29
40 and above	432	49.71	432	52.57	456	1.58	492	30.23	493	128.82

[4] Westfall PH, Young SS. Resampling-based multiple testing: Examples and methods for p-value adjustment. Wiley & Sons, 1993; Kling JR, Liebman JB, Katz LF. Experimental Analysis of Neighborhood Effects. Econometrica. 2007;75(1):83-119; Jones D, Molitor D, Reif J. What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study. NBER Working Paper No. 24229. 2018.

Notes: N denotes control group sample sizes only. Control means are weighted by the combination of a weight for exposure to the wellness program and a weight that balances treatment and control samples on demographics.

Equation 4 shows this interaction in our base ITT framework for age (characterized by an indicator for being age 40 or over).

$$Y_{ij} = \beta_0 + \beta_1 TREATMENT_j + \beta_2 Age40_i * TREATMENT_j + \beta_3 \boldsymbol{X}_{ij} + \varepsilon_{ij} \quad (4)$$

The effect of the wellness program on those under 40 is estimated by β_1, while the effect for those 40 and older is estimated by the sum of the coefficients $\beta_1 + \beta_2$. Age categories continue to be included in covariates $\boldsymbol{X}$.

E. Worksite-level analyses

We complemented our analyses at the individual level with analyses at the worksite level. Worksite-level data were generated by aggregating employees to worksites based on the worksite location of their first appearance in the data. We regression-adjusted each outcome for demographics at the individual level before aggregation, and weighted individuals based on their hours worked to calculate worksite-level averages (which reflect an average hour worked at the worksite). For analyses of medical and pharmaceutical outcomes that involved counts (either dollars or utilization), we multiplied worksite-level outcomes by 2000 to allow the data to approximate the magnitude of a full-time equivalent employee (2000 hours). The resulting worksite-level dataset for each outcome comprised 160 data points, one for each worksite (20 intervention, 20 primary control, and 120 secondary control).

We focused on outcomes measured in administrative data for all employees, dictated by data availability but also representing the employer perspective on aggregate outcomes affected by the decision to have a wellness program. Our estimation equation is of the form:

$$Y'_j = \beta_0 + \beta_1 TREATMENT_j + \varepsilon_j \quad (5)$$

In equation (5), the subscript j denotes a worksite. Y'_j represents a worksite-level average outcome. $TREATMENT_j$ is a binary indicator of randomization into the treatment group, with β_1 indicating the average site-level effect of being randomized into treatment. Covariates were, as noted, incorporated at the individual level before aggregation. Standard errors were adjusted for heteroscedasticity.

F. Sensitivity analyses and secondary analyses

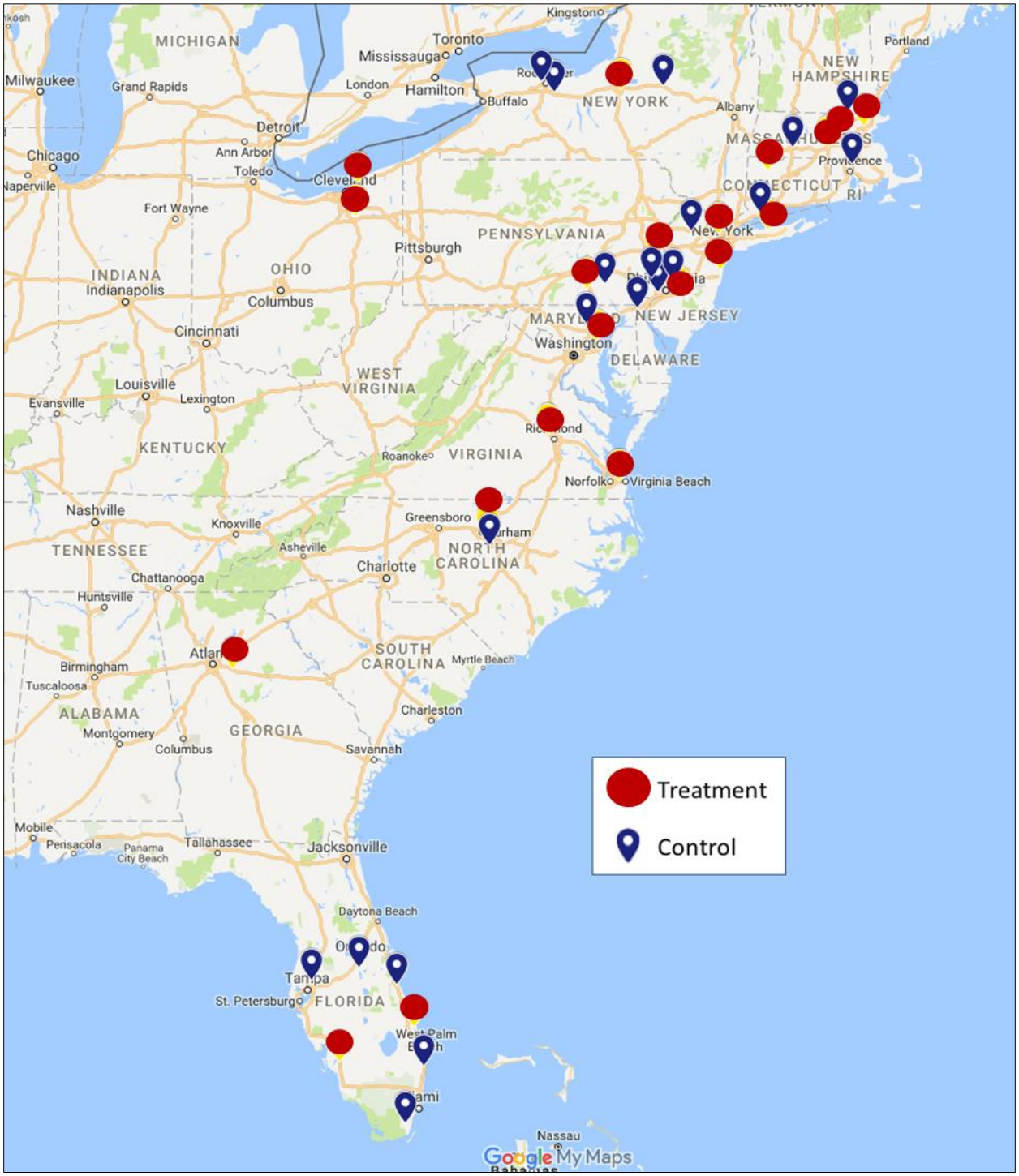
In the statistical analyses above, our base regression models use least squares specifications (OLS for ITT, 2SLS for LATE) for both continuous and binary outcomes. This approach has both strengths and weaknesses.[5] To test the robustness of our results, we estimated alternative functional forms, notably logit models for binary outcome variables.

[5] See, for example, Buntin MB, Zaslavsky AM. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ. 2004 May;23(3):525-42; Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ. 2005 May;24(3):465-88.

In secondary analyses, we also reassessed a set of key outcomes using the exposure weights without the balance weights. The key outcomes were: total medical spending, total prescription drug spending, absenteeism, systolic blood pressure, BMI, annual exam (binary), SF-8 mental and physical health score, regular exercise, number of sweetened drinks per day, smoking status (binary), and the number of alcoholic drinks consumed per week.

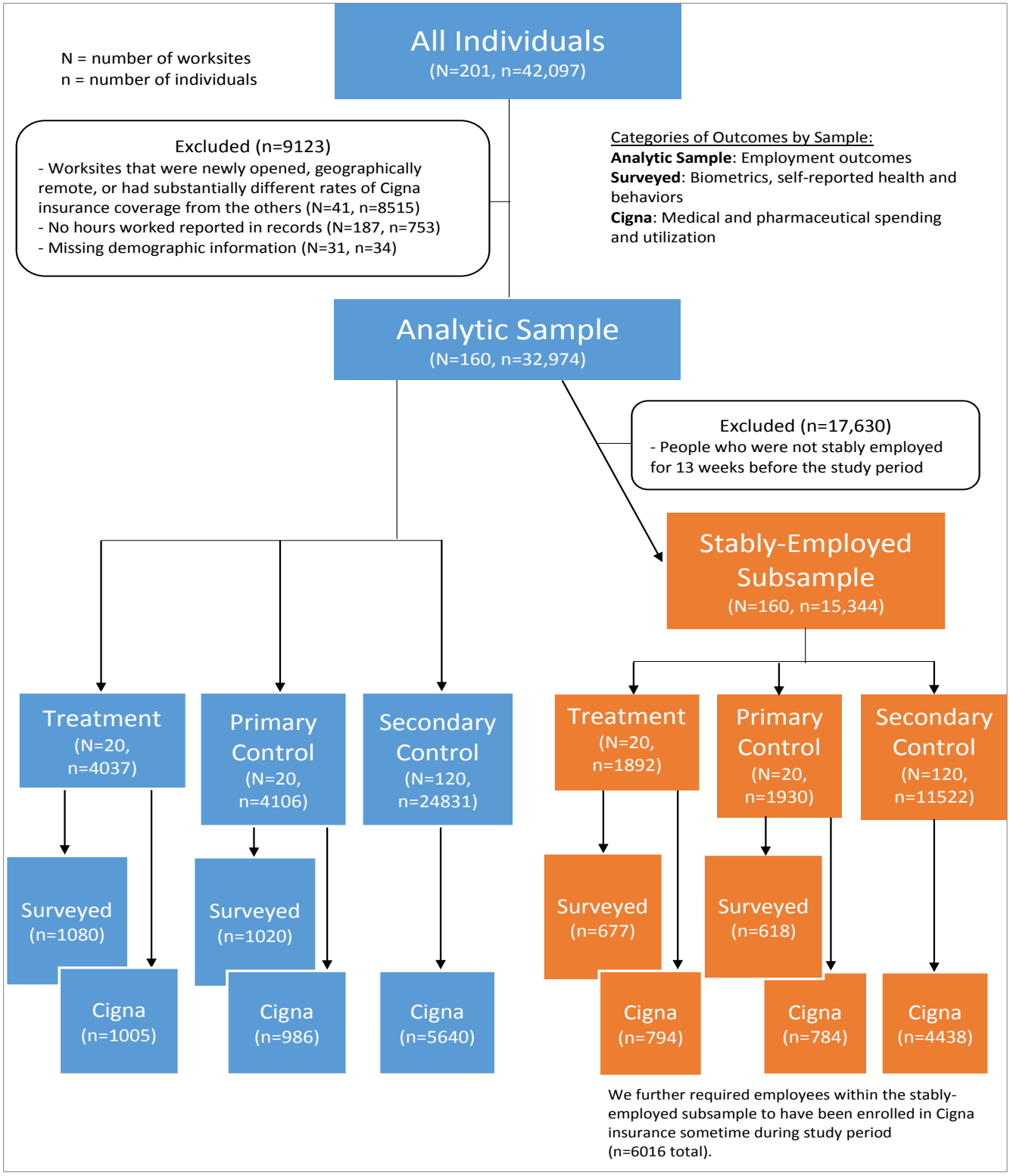
Given uncertainty about the functional form of the effect of participation in a multifaceted program on outcomes, we tested the sensitivity of our results using alternative definitions of participation. As noted above in the first stage estimates, we tested a definition of participation based on a threshold of completing at least 3 modules, as well as a continuous metric of participation as the number of modules completed. We presented the results of these sensitivity and secondary analyses in additional tables that allow comparison to the main estimates.

eFigure 1: Location of Treatment and Control Worksites



Notes: This map shows the randomly selected 20 treatment and 20 primary control worksites within BJ's Wholesale Club in phase 1 of this workplace wellness program. Red markers designate treatment worksites. Blue markers designate primary control worksites.

eFigure 2. Treatment Assignment and Analytic Samples



Notes: The vast majority (97.4%) of employees worked at one worksite during the study period; the remaining employees who worked at more than one site were analyzed according to their initial site of random assignment.

eTable 1. Timeline of the Workplace Wellness Intervention

	Program announced ↓					**Registered Dietitians begin working in the treatment sites ↓**														
Year	**2015**												**2016**							
Month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Program Announcement																				
Module 1. Take Charge of Your Health																				
Module 2. Nutrition for a Lifetime																				
Module 3. Club Cardio Challenge Round 1																				
Module 4. Club Cardio Challenge Round 2																				
Module 5. Maintain Don't Gain																				
Module 6. Power Down the Pressure																				
Module 7. Weight Loss Boot Camp																				
Module 8. Movin' in May																				
Primary Data Collection																				

Notes: This table presents a graphical illustration of the timing the workplace wellness intervention. The intervention began in January 2015 and comprised 8 wellness program modules, the last of which concluded at the end of June 2016. Following its conclusion, personal health assessments and in-person screenings were conducted during the summer of 2016 (primary data collection).

eTable 2. Summary of Outcome Data Collected

Data	Data Availability		
	Treatment Worksites (20 worksites)	**Primary Control Worksites (20 worksites)**	**Secondary Control Worksites (120 worksites)**
In-Person Data			
Clinical Measures of Health (Screening)	Employees completing screening	Employees completing screening	None (by design)
Personal Health Assessment (Survey)	Employees completing survey	Employees completing survey	None (by design)
Administrative Data			
Employment records	All employees	All employees	All employees
Claims data (medical and pharmaceutical)	Employees insured by Cigna	Employees insured by Cigna	Employees insured by Cigna

Notes: In-person data consisted of biometric screening data collected during in-person screenings conducted by registered nurses and self-reported data gathered from concurrently administered personal health assessment surveys. The biometric data included blood pressure, height and weight (allowing for the calculation of BMI), and blood measurements of cholesterol and blood sugar. Personal health assessments included self-reported information on health behaviors, health, and wellbeing. These primary data were available for the individuals in the 20 treatment worksites and the 20 primary control worksites who completed the screenings after the 18-month intervention period. For completing this primary data collection, an employee received a $50 gift card. The participation rate in this data collection was about 52% in treatment worksites and 49% in control worksites. Administrative data included employment records and claims data, both of which were available for all employees across treatment and control worksites. Employment records included data on employment history and earnings. Medical and pharmaceutical claims data were available through Cigna for the subset of employees who were insured through a Cigna plan, the large majority of whom were full-time employees. BJ's Wholesale Club is a self-insured company (i.e. it bears risk for the health care spending of its employed population), with Cigna as the administrator of its health plans. At any given time during the study period, about 35% of employees in the company had employer-sponsored health insurance through Cigna.

eTable 3. Pre-Specified Outcomes

(1) Self-Reported Health and Behaviors

- Annual exam (%)
- Flu shot (%)
- Percent of other recommended tests received
- PHQ-2 score of 3 or above (%)
- SF-8 score – physical summary score
- SF-8 score – mental summary score
- Unmanaged stress (%)
- Stress at work (%)
- Good quality, adequate amount of sleep (%)
- Regular exercise (%)
- Three or more days of moderate exercise per week (%)
- Number of days per week intentionally increase activity
- Number of hours sitting per day
- Number of meals eaten out
- Number of naturally or artificially sweetened drinks per day
- Read the Nutrition Facts panel (%)
- Consume at least 2 cups of fruit and 2.5 cups of vegetables per day (%)
- Choose whole grain foods and reduced fat foods more often than the regular variety (%)
- Considering losing weight in the next 6 months (%)
- Actively managing weight (%)
- Smoker (%)
- Number of drinks per week
- Number doctor visits in last 12 months
- Any doctor visit in last 12 months (%)
- Any ER visit in last 12 months (%)
- Ever hospital patient in the last 12 months (%)
- Days spent in hospital
- Number different prescriptions last 12 months
- Any prescriptions in last 12 months (%)

(2) Clinical Measures of Health

- Cholesterol
- HDL
- Glucose
- Systolic BP
- Diastolic BP
- BMI
- High Cholesterol (Cholesterol >= 200)
- Low HDL (HDL < 40)
- Hypertensive (Systolic BP >= 140 or Diastolic >= 90)
- Obese (BMI >=30)

(3) Health Care Spending and Utilization

- Medical Spending
  - Total Spending
  - Out-of-pocket Spending
  - Office

Inpatient Hospital
Emergency Room
Urgent Care
Other
Medical Utilization
Any Physician Visit (%)
Number of Physician Visits
Any Hospitalization (%)
Number of Hospitalizations
Any ER Visit (%)
Number of ER Visits
Any Urgent Care Visit (%)
Number of Urgent Care Visits
Any Preventive Care Visit (%)
Number of Preventive Care Visits
Pharmaceutical Spending
Total Spending
Out-of-pocket Spending
Pharmaceutical Utilization
Any Medications (%)
Number of Distinct Medications
Number of Medication Months
Any Asthma Medications (%)
Number of Asthma Medication Months
Any Cardiovascular Medications (%)
Number of Cardiovascular Medication Months
Any Diabetes Medications (%)
Number of Diabetes Medication Months
Any Hyperlipidemia Medications (%)
Number of Hyperlipidemia Medication Months
Any Mental Health Medications (%)
Number of Mental Health Medication Months
Any Pain Medications (%)
Number of Pain Medication Months
Any Antibiotics Medications (%)
Number of Antibiotics Medication Months
Any Other Medications (%)
Number of Other Medication Months
(4) Employment Outcomes
Absenteeism (%)
Performance Review (% with 3 or better)
Tenure

Notes: This table provides the complete list of pre-specified outcomes analyzed in this clinical trial. They were pre-specified in the publicly archived Analysis Plan: Song Z, Baicker K. The Impact of Workplace Wellness on Health, Health Care, and Employment Outcomes: A Randomized Controlled Trial, Analysis Plan—Phase 1. 2018 Mar 15. (https://clinicaltrials.gov/ProvidedDocs/58/NCT03167658/Prot_SAP_000.pdf); Baicker K, Maylone B, Song Z. The Impact of Employee Wellness Programs: A Randomized Controlled Trial. American Economic Association RCT Registry. 2018 Mar 15. (https://www.socialscienceregistry.org/trials/586).

eTable 4. Characteristics of the Study Population

A: All Employees

	All employees at Treatment Worksites (n=4,037)	All employees at Primary Control Worksites (n=4,106)	All employees at Primary & Secondary Control Worksites (n=28,937)
Demographic Characteristics			
Age (yrs)	38.76	38.32	38.66
Female (%)	46.33	45.46	45.82
Race (%)			
Black	19.80	20.08	20.67
White	56.29	57.88	55.28
Hispanic	17.87	17.06	17.84
Other	6.03	4.98	6.21

Notes: All individuals were assigned to treatment or control status based on the first worksite in which they appeared during the treatment period. Characteristics were weighted by exposure to the wellness program based on duration of employment. Age was defined as of December, 2014, the year pre-intervention. Demographic characteristics were measured pre-intervention.

B: Participants in the Personal Health Assessment

	All employees at Treatment Worksites (n=1,080)	All employees at Primary Control Worksites (n=1,020)
Demographic Characteristics		
Age (yrs)	41.23	40.08
Female (%)	57.56	57.12
Race (%)		
Black	19.00	19.11
White	57.50	58.96
Hispanic	18.56	16.76
Other	4.94	5.16

Notes: The Personal Health Assessment (survey) was offered only at the 20 treatment worksites and 20 primary control worksites. Employees were included if they answered at least 1 question on the PHA. All individuals were assigned to treatment or control status based on the first worksite in which they appeared during the treatment period. Characteristics were weighted by exposure to the wellness program based on duration of employment and a weight that balances treatment and control on demographics. Age was defined as of December, 2014, the year pre-intervention. Demographic characteristics were measured pre-intervention.

C: Subsample with Employer-Sponsored Insurance (Cigna)

	All employees at Treatment Worksites (n=1,005)	All employees at Primary Control Worksites (n=986)	All employees at Primary & Secondary Control Worksites (n= 6,626)
Demographic Characteristics			
Age (yrs)	44.85	43.83	44.26
Female (%)	46.99	44.76	46.61
Race (%)			
Black	15.96	15.31	16.94
White	60.83	66.11	62.04
Hispanic	17.93	14.51	15.97
Other	5.28	4.07	5.05

Notes: Employees were included if they had at least 1 month of employer-sponsored insurance (Cigna) through the employer. All individuals were assigned to treatment or control status based on the first worksite in which they appeared during the study period. Characteristics were weighted by exposure to the wellness program based on duration of employment and a weight that balances treatment and control on demographics. Age was defined as of December, 2014, the year pre-intervention. Demographic characteristics were measured pre-intervention.

D: Stably Employed Subsample

	Stably Employed Subsample at Treatment Worksites (n=1,892)	Stably Employed Subsample at Primary Control Worksites (n=1,930)	Stably Employed Subsample at Primary & Secondary Control Worksites (n=13,452)
Demographic Characteristics			
Age (yrs)	41.06	39.99	40.64
Female (%)	46.73	45.29	46.03
Race (%)			
Black	18.79	18.58	19.09
White	57.05	59.82	56.99
Hispanic	18.52	17.02	17.89
Other	5.65	4.57	6.03

Notes: Employees are included if they were part of the stably employed subsample. All individuals were assigned to treatment or control status based on the first worksite in which they appeared during the study period. Characteristics were weighted by exposure to the wellness program based on duration of employment and a weight that balances treatment and control on demographics. Age was defined as of December, 2014, the year pre-intervention. Demographic characteristics were measured pre-intervention.

E: Worksite Level

	Treatment Worksites (n=20)	Primary Control Worksites (n=20)	Primary & Secondary Control Worksites (n=140)
Employee Demographics			
Age (yrs)	39.44	38.13	38.42
Female (%)	48.85	46.01	45.77
Race (%)			
Black	18.04	20.33	19.99
White	68.05	59.45	57.81
Hispanic	9.04	16.62	16.58
Other	4.87	3.60	5.61
Area Demographics			
Age (yrs)	40.75	39.49	39.6
Female (%)	51.23	51.40	51.3
Race (%)			
Black	12.19	13.82	13.30
White	77.74	76.32	74.48
Hispanic	9.36	15.41	13.80
Other	10.07	9.84	12.20

Notes: This table shows characteristics at the worksite level. Area demographics are taken from the 2015 American Community Survey (ACS) Population Estimates for the county each worksite was located in. Worksite-level analyses were obtained by first calculating a weighted average for each worksite (weighted by an employee's hours worked during the study period). Treatment status was defined by the first worksite an employee appeared in during the study period. Age was defined as of December, 2014, the year pre-intervention. Demographic characteristics were measured pre-intervention.

eTable 5: Average Participation Rates by Module

A. Participation by Module (%)

	Take Charge of Your Health	Nutrition for a Lifetime	Club Cardio Challenge Round 1	Club Cardio Challenge Round 2	Maintain Don't Gain	Power Down the Pressure	Weight Loss Boot Camp	Movin' in May
Mean	12.2	25.6	37.7	28.6	31.6	33.4	28.7	28.5
Median	8.8	23.5	30.6	22.6	32.4	33.1	29.7	27.6
Minimum	0.0	13.2	19.8	16.2	14.0	14.7	9.0	13.7
Maximum	32.4	42.6	70.0	67.2	65.5	56.4	55.1	50.9

Notes: The participation rate is calculated as the number of individuals who completed a module divided by the number of employees eligible to complete that module when it was running in the treatment worksite. Means were weighted by program exposure across all treatment worksites, based on employees' number of employment days during a module. The median, minimum, and maximum participation rates were calculated at the worksite level among the sample of 20 worksites. For descriptions of the modules, please refer to eMethods 1.

B. Participation in the Wellness Program and Among Participants

Definition of Participation	Employees at Treatment Worksites (N= 4,037)	Extent of Participation Among Employees Participating in at Least One Module at Treatment Worksites (N= 1,421)
Completed any module (%)	35.2	100.0
Modules completed (#)	1.3	3.7
Completed 3 or more modules (%)	21.4	60.9

Notes: This table presents average participation rates calculated using different definitions of participation.

eTable 6. Mean Values and Effect of Program on Self-Reported Health and Behaviors

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
Screenings and Exams						
Annual exam (%)	65.51 (47.56)	-1.26 -6.96 - 4.45 (2.87) [0.66] {1.00}	-1.61 -8.70 - 5.49 (3.62) [0.66] {1.00}	65.10 (47.71)	-1.33 -7.68 - 5.01 (3.19) [0.68] {1.00}	-1.65 -9.24 - 5.95 (3.87) [0.67] {1.00}
Flu shot (%)	35.16 (47.77)	-2.42 -8.31 - 3.47 (2.96) [0.42] {1.00}	-3.07 -10.38 - 4.23 (3.73) [0.41] {1.00}	33.42 (47.21)	-2.30 -8.20 - 3.60 (2.96) [0.44] {1.00}	-2.81 -9.84 - 4.22 (3.59) [0.43] {1.00}
Percent of other recommended tests received	55.93 (31.01)	3.18 0.02 - 6.35 (1.59) [0.05] {0.69}	4.07 0.07 - 8.07 (2.04) [0.05] {0.71}	57.09 (30.57)	3.40 0.03 - 6.78 (1.70) [0.05] {0.66}	4.18 0.11 - 8.25 (2.08) [0.04] {0.69}
Mental Health and Well-being						
PHQ-2 score of 3 or above (%)	8.55 (27.97)	-0.97 -3.45 - 1.51 (1.25) [0.44] {1.00}	-1.24 -4.32 - 1.84 (1.57) [0.43] {1.00}	8.40 (27.76)	-0.82 -3.80 - 2.15 (1.49) [0.58] {1.00}	-1.01 -4.53 - 2.51 (1.79) [0.57] {1.00}
SF-8 score – physical summary score	50.79 (7.71)	-0.15 -0.84 - 0.54 (0.35) [0.66] {1.00}	-0.19 -1.04 - 0.66 (0.44) [0.66] {1.00}	50.92 (7.71)	-0.34 -1.11 - 0.44 (0.39) [0.39] {1.00}	-0.41 -1.32 - 0.50 (0.46) [0.38] {1.00}
SF-8 score – mental summary score	51.17 (9.08)	-0.35 -1.24 - 0.55 (0.45) [0.44] {1.00}	-0.44 -1.55 - 0.67 (0.57) [0.43] {1.00}	51.21 (9.08)	-0.38 -1.34 - 0.58 (0.48) [0.43] {1.00}	-0.46 -1.59 - 0.67 (0.58) [0.42] {1.00}
Unmanaged stress (%)	41.77 (49.35)	-2.71 -7.68 - 2.25 (2.49) [0.28] {0.99}	-3.47 -9.65 - 2.70 (3.15) [0.27] {0.99}	41.38 (49.30)	-2.80 -8.47 - 2.86 (2.85) [0.33] {0.99}	-3.44 -10.19 - 3.31 (3.44) [0.32] {0.99}
Stress at work (%)	55.70 (49.70)	1.98 -2.63 - 6.59 (2.32) [0.40] {1.00}	2.53 -3.21 - 8.27 (2.93) [0.39] {1.00}	58.24 (49.36)	3.01 -1.83 - 7.86 (2.43) [0.22] {0.98}	3.70 -2.06 - 9.46 (2.94) [0.21] {0.98}
Sleep						
Good quality, adequate amount of sleep (%)	54.07 (49.86)	-2.08 -5.99 - 1.84 (1.97) [0.29] {0.99}	-2.66 -7.55 - 2.23 (2.49) [0.29] {0.99}	54.43 (49.84)	-2.73 -7.49 - 2.04 (2.39) [0.26] {0.99}	-3.36 -9.03 - 2.32 (2.90) [0.25] {0.98}
Physical Activity						
Regular exercise (%)	61.93 (48.58)	8.33 3.86 - 12.79 (2.25) [0.00] {0.03}	10.64 5.26 - 16.03 (2.75) [0.00] {0.03}	63.25 (48.25)	8.45 3.60 - 13.3 (2.44) [0.00] {0.06}	10.36 4.79 - 15.93 (2.84) [0.00] {0.05}

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
per week (%)	63.97 (48.04)	4.11 -0.59 - 8.80 (2.36) [0.09] {0.85}	5.27 -0.57 - 11.12 (2.98) [0.08] {0.84}	64.12 (48.01)	3.49 -1.71 - 8.68 (2.61) [0.19] {0.97}	4.30 -1.90 - 10.49 (3.16) [0.17] {0.97}
Number of days per week intentionally increase activity	3.05 (2.37)	0.08 -0.13 - 0.30 (0.11) [0.44] {1.00}	0.11 -0.16 - 0.38 (0.14) [0.44] {1.00}	3.06 (2.36)	0.17 -0.08 - 0.42 (0.12) [0.17] {0.97}	0.21 -0.09 - 0.50 (0.15) [0.17] {0.97}
Number of hours sitting per day	3.50 (1.73)	0.02 -0.18 - 0.22 (0.10) [0.83] {1.00}	0.03 -0.22 - 0.28 (0.13) [0.83] {1.00}	3.50 (1.73)	0.06 -0.13 - 0.26 (0.10) [0.52] {1.00}	0.08 -0.16 - 0.31 (0.12) [0.51] {1.00}
Nutrition						
Number of meals eaten out	1.85 (1.57)	-0.06 -0.21 - 0.10 (0.08) [0.48] {1.00}	-0.07 -0.26 - 0.12 (0.10) [0.47] {1.00}	1.82 (1.54)	-0.01 -0.18 - 0.15 (0.08) [0.89] {1.00}	-0.01 -0.21 - 0.18 (0.10) [0.89] {1.00}
Number of naturally or artificially sweetened drinks per day	1.85 (1.86)	0.10 -0.11 - 0.32 (0.11) [0.34] {1.00}	0.13 -0.13 - 0.40 (0.14) [0.33] {1.00}	1.81 (1.84)	0.10 -0.13 - 0.34 (0.12) [0.39] {0.99}	0.13 -0.16 - 0.42 (0.15) [0.38] {1.00}
Read the Nutrition Facts panel (%)	58.65 (49.27)	4.37 -1.06 - 9.80 (2.73) [0.11] {0.91}	5.57 -1.12 - 12.27 (3.42) [0.10] {0.91}	58.38 (49.33)	7.86 1.52 - 14.19 (3.18) [0.02] {0.37}	9.62 2.20 - 17.04 (3.79) [0.01] {0.33}
Consume at least 2 cups of fruit and 2.5 cups of vegetables per day (%)	57.54 (49.45)	3.29 -1.07 - 7.66 (2.20) [0.14] {0.93}	4.22 -1.19 - 9.63 (2.76) [0.13] {0.92}	57.29 (49.50)	4.76 -0.48 - 9.99 (2.63) [0.07] {0.79}	5.86 -0.33 - 12.05 (3.16) [0.06] {0.78}
Choose whole grain foods and reduced fat foods more often than the regular variety (%)	33.17 (47.11)	1.22 -3.17 - 5.61 (2.21) [0.58] {1.00}	1.55 -3.89 - 6.99 (2.78) [0.58] {1.00}	34.53 (47.58)	1.49 -3.68 - 6.65 (2.60) [0.57] {1.00}	1.82 -4.32 - 7.95 (3.13) [0.56] {1.00}
Weight Management						
Considering losing weight in the next 6 months (%)	56.26 (49.63)	9.54 3.71 - 15.36 (2.93) [0.00] {0.09}	12.06 4.77 - 19.35 (3.72) [0.00] {0.11}	56.45 (49.63)	9.36 3.60 - 15.12 (2.89) [0.00] {0.10}	11.36 4.46 - 18.26 (3.52) [0.00] {0.11}
Actively managing weight (%)	54.70 (49.81)	13.61 7.06 - 20.16 (3.29) [0.00] {0.02}	17.24 9.12 - 25.35 (4.14) [0.00] {0.01}	54.74 (49.82)	13.86 7.04 - 20.68 (3.42) [0.00] {0.02}	16.84 8.81 - 24.87 (4.10) [0.00] {0.02}
Tobacco Use						

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
Smoker (%)	24.63 (43.11)	-6.87 -12.85 - -0.88 (3.01) [0.03] {0.52}	-8.78 -16.30 - -1.26 (3.84) [0.02] {0.53}	24.57 (43.08)	-8.88 -15.43 - -2.33 (3.29) [0.01] {0.27}	-10.91 -18.79 - -3.02 (4.02) [0.01] {0.28}
Alcohol Use						
Number of drinks per week	4.65 (7.41)	-0.57 -1.13 - -0.01 (0.28) [0.04] {1.00}	-0.73 -1.43 - -0.03 (0.36) [0.04] {0.68}	4.72 (7.39)	-0.36 -1.04 - 0.32 (0.34) [0.30] {0.99}	-0.44 -1.25 - 0.37 (0.41) [0.29] {0.99}
Medical Utilization						
Number doctor visits in last 12 months	1.52 (1.12)	0.00 -0.12 - 0.12 (0.06) [0.98] {1.00}	0.00 -0.14 - 0.15 (0.07) [0.98] {1.00}	1.53 (1.11)	0.00 -0.13 - 0.13 (0.07) [0.98] {1.00}	0.00 -0.15 - 0.16 (0.08) [0.98] {1.00}
Any doctor visit in last 12 months (%)	75.55 (43.00)	-0.59 -5.30 - 4.13 (2.37) [0.80] {1.00}	-0.75 -6.59 - 5.10 (2.98) [0.80] {1.00}	76.06 (42.70)	-0.45 -6.02 - 5.13 (2.80) [0.87] {1.00}	-0.54 -7.14 - 6.05 (3.36) [0.87] {1.00}
Any ER visit in last 12 months (%)	25.84 (43.80)	-3.51 -8.03 - 1.01 (2.27) [0.13] {0.92}	-4.47 -10.10 - 1.17 (2.87) [0.12] {0.92}	25.31 (43.51)	-3.52 -8.59 - 1.54 (2.54) [0.17] {0.97}	-4.29 -10.31 - 1.73 (3.07) [0.16] {0.97}
Ever hospital patient in the last 12 months (%)	17.52 (38.04)	-2.94 -6.95 - 1.08 (2.02) [0.15] {0.93}	-3.69 -8.63 - 1.26 (2.52) [0.14] {0.93}	17.53 (38.06)	-3.25 -8.12 - 1.62 (2.45) [0.19] {0.97}	-3.91 -9.63 - 1.81 (2.92) [0.18] {0.97}
Days spent in hospital	0.43 (1.37)	-0.08 -0.24 - 0.07 (0.08) [0.28] {0.99}	-0.11 -0.29 - 0.08 (0.10) [0.27] {0.99}	0.43 (1.40)	-0.13 -0.30 - 0.04 (0.08) [0.13] {0.93}	-0.16 -0.35 - 0.04 (0.10) [0.12] {0.93}
Number of different prescriptions in the last 12 months	1.32 (1.64)	-0.06 -0.23 - 0.10 (0.08) [0.43] {1.00}	-0.08 -0.28 - 0.12 (0.10) [0.43] {1.00}	1.30 (1.64)	-0.04 -0.21 - 0.14 (0.09) [0.70] {1.00}	-0.04 -0.25 - 0.17 (0.11) [0.69] {1.00}
Any prescriptions in last 12 months (%)	52.81 (49.95)	-1.76 -5.99 - 2.46 (2.12) [0.41] {1.00}	-2.25 -7.51 - 3.01 (2.68) [0.40] {1.00}	52.33 (49.99)	-0.78 -5.46 - 3.90 (2.35) [0.74] {1.00}	-0.95 -6.49 - 4.60 (2.83) [0.74] {1.00}
Standardized treatment effect (mental health and well-being)		0.00 -0.05 - 0.05 (0.03) [0.97]	0.00 -0.06 - 0.07 (0.04) [0.97]		-0.01 -0.07 - 0.04 (0.03) [0.66]	-0.02 -0.08 - 0.05 (0.04) [0.66]
Standardized treatment effect (health behaviors)		0.07 0.02 - 0.10	0.09 0.03 - 0.13		0.07 0.02 - 0.11	0.09 0.03 - 0.13

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
		(0.02) [0.00]	(0.03) [0.00]		(0.02) [0.00]	(0.03) [0.00]
N	858 - 1009	1722 - 2022		536 - 641	1078 - 1291	

Notes: This table reports the coefficient on TREATMENT from estimating equation (1) by OLS (column 2), and the coefficient on PARTICIPATION from estimating equation (2) by IV (column 3). Standard errors are shown in parentheses with p-values in brackets and family-wise p-values in curly braces. Column 1 reports the mean of each self-reported health outcome in the control group for each sample (with standard deviation in parentheses). All regressions included demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, Cigna coverage status, full-time status, paid hourly status, and job category) and clustered standard errors by the worksite. The control means and regressions were weighted by the combination of a weight for exposure to the wellness program and a weight that balances treatment and control samples on demographics. The standardized treatment effect for mental health and well-being was calculated using the outcomes under the Mental Health and Well-being section, and the standardized treatment effect for health behaviors was calculated using the outcomes under the Screenings and Exams, Sleep, Physical Activity, Nutrition, Weight Management, Tobacco Use, and Alcohol Use sections. Sample sizes for this domain of outcomes differed across outcomes based on the number of respondents for each outcome. Thus, this table reports the range of values for sample size

eTable 7. Mean Values and Effect of Program on Clinical Measures of Health

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
Continuous Measures						
Total Cholesterol (mg/dl)	177.55 (41.48)	2.61 -5.75 - 10.98 (4.21) [0.54] {0.99}	3.35 -7.11 - 13.81 (5.34) [0.53] {0.99}	178.83 (41.40)	2.72 -6.47 - 11.91 (4.62) [0.56] {0.99}	3.34 -7.64 - 14.33 (5.61) [0.55] {0.99}
High density lipoprotein (HDL) cholesterol (mg/dl)	52.96 (16.36)	-0.33 -2.39 - 1.73 (1.03) [0.75] {1.00}	-0.43 -3.01 - 2.16 (1.32) [0.75] {1.00}	53.54 (16.39)	-0.50 -2.76 - 1.76 (1.14) [0.66] {0.99}	-0.61 -3.31 - 2.09 (1.38) [0.66] {0.99}
Glucose (mg/dl)	101.94 (33.46)	1.40 -4.04 - 6.84 (2.74) [0.61] {1.00}	1.79 -5.02 - 8.61 (3.48) [0.61] {1.00}	101.73 (32.02)	2.11 -3.69 - 7.91 (2.92) [0.47] {0.98}	2.59 -4.35 - 9.53 (3.54) [0.46] {0.99}
Blood pressure, systolic (mm Hg)	124.30 (16.88)	0.24 -1.68 - 2.15 (0.96) [0.81] {1.00}	0.30 -2.09 - 2.70 (1.22) [0.80] {1.00}	124.86 (16.80)	0.29 -1.72 - 2.30 (1.01) [0.77] {0.99}	0.36 -2.03 - 2.76 (1.22) [0.77] {0.99}
Blood pressure, diastolic (mm Hg)	79.70 (10.57)	0.49 -0.81 - 1.78 (0.65) [0.46] {0.98}	0.62 -0.99 - 2.24 (0.82) [0.45] {0.98}	80.10 (10.50)	0.55 -0.88 - 1.97 (0.72) [0.45] {0.98}	0.67 -1.03 - 2.37 (0.87) [0.44] {0.99}
Body mass index (BMI)	29.70 (7.09)	0.09 -0.59 - 0.78 (0.34) [0.79] {1.00}	0.12 -0.74 - 0.98 (0.44) [0.78] {1.00}	29.61 (7.00)	0.23 -0.51 - 0.97 (0.37) [0.54] {0.99}	0.28 -0.60 - 1.16 (0.45) [0.53] {0.99}
Binary Measures (%)						
High total cholesterol (total cholesterol ≥200 mg/dl)	29.34 (45.55)	0.07 -7.97 - 8.11 (4.04) [0.99] {1.00}	0.09 -9.95 - 10.14 (5.13) [0.99] {1.00}	30.68 (46.15)	-0.64 -9.79 - 8.51 (4.60) [0.89] {0.99}	-0.79 -11.70 - 10.12 (5.57) [0.89] {0.99}
Low HDL cholesterol (HDL cholesterol <40 mg/dl)	22.32 (41.66)	-1.10 -5.83 - 3.64 (2.38) [0.65] {1.00}	-1.41 -7.36 - 4.53 (3.03) [0.64] {1.00}	20.81 (40.63)	-0.12 -5.46 - 5.22 (2.68) [0.96] {0.99}	-0.15 -6.52 - 6.23 (3.25) [0.96] {0.99}
Hypertension (systolic BP ≥140 or diastolic BP ≥90 mm Hg)	23.13 (42.18)	2.68 -2.43 - 7.79 (2.57) [0.30] {0.93}	3.43 -2.93 - 9.80 (3.25) [0.29] {0.92}	24.20 (42.86)	3.29 -2.47 - 9.04 (2.90) [0.26] {0.89}	4.03 -2.82 - 10.87 (3.49) [0.25] {0.89}

	All Employees			Stably Employed Subsample		
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect
	(1)	(2)	(3)	(1)	(2)	(3)
Obesity (BMI ≥30)	43.00 (49.53)	0.55 -3.66 - 4.76 (2.12) [0.80] {1.00}	0.71 -4.57 - 5.99 (2.69) [0.79] {1.00}	42.67 (49.50)	1.23 -3.61 - 6.06 (2.43) [0.62] {0.99}	1.51 -4.28 - 7.29 (2.95) [0.61] {0.99}
Standardized treatment effect (clinical health outcomes)		-0.03 -0.09 - 0.03 (0.03) [0.37]	-0.04 -0.12 - 0.04 (0.04) [0.36]		-0.04 -0.11 - 0.02 (0.04) [0.22]	-0.05 -0.13 - 0.03 (0.04) [0.21]
N	1046 - 1074	2082 - 2139		658 - 671	1311 - 1338	

Notes: This table reports the coefficient on TREATMENT from estimating equation (1) by OLS (column 2), and the coefficient on PARTICIPATION from estimating equation (2) by IV (column 3). Standard errors are listed in parentheses with p-values in brackets and family-wise p-values in curly braces. Column 1 reports the mean of each biometric outcome in the control group for each sample (with standard deviation in parentheses). All regressions included demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, Cigna coverage status, full-time status, paid hourly status, and job category) and clustered standard errors by the worksite. The control means and regressions were weighted by the combination of a weight for exposure to the wellness program and a weight that balances treatment and control samples on demographics. The standardized treatment effect for this domain of outcomes was calculated using only the continuous outcome variables. Sample sizes for this domain differed across outcomes based on the number of data participants for each outcome. Thus, this table reports the range of values for sample size.

eTable 8. Mean Values and Effect of Program on Medical Spending and Utilization

	Employee-level			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)
Spending								
Total Spending	3953.01 (14697.45)	-425.57 -1266 - 415 (425.73) [0.32] {0.95}	-670.13 -1954 - 614 (655.05) [0.31] {0.95}	3810.21 (12972.22)	-466.50 -1321 - 388 (432.56) [0.28] {0.95}	-724.80 -2012 - 562 (656.71) [0.27] {0.95}	4911.45 (3137.26)	-712.00 -1828 - 404 (565.10) [0.21] {0.90}
Out-of-pocket Spending	778.49 (1208.34)	-7.93 -113 - 97 (53.03) [0.88] {1.00}	-12.49 -175 - 150 (83.13) [0.88] {1.00}	740.54 (1053.86)	-23.51 -95 - 48 (36.44) [0.52] {0.99}	-36.52 -147 - 74 (56.23) [0.52] {0.99}	851.11 (242.31)	-33.01 -138 - 72 (53.19) [0.54] {0.99}
By Site of Care:								
Office	2132.87 (7361.56)	-222.01 -723 - 278 (253.41) [0.38] {0.97}	-349.59 -1119 - 419 (392.34) [0.37] {0.96}	2166.29 (7458.85)	-230.50 -815 - 354 (295.91) [0.44] {0.99}	-358.13 -1245 - 529 (452.69) [0.43] {0.98}	2592.05 (1679.25)	-269.42 -857 - 318 (297.54) [0.37] {0.95}
Inpatient Hospital	1150.87 (9227.73)	-234.10 -706 - 238 (238.88) [0.33] {0.96}	-368.63 -1092 - 355 (369.09) [0.32] {0.95}	1008.72 (7289.77)	-218.86 -648 - 210 (217.35) [0.32] {0.96}	-340.05 -989 - 309 (331.02) [0.30] {0.96}	1535.28 (1645.66)	-420.63 -1131 - 290 (359.56) [0.24] {0.92}
Emergency Room	526.84 (1750.45)	78.49 -103 - 260 (91.77) [0.39] {0.97}	123.60 -159 - 407 (144.38) [0.39] {0.97}	494.83 (1606.94)	64.09 -76 - 205 (71.14) [0.37] {0.98}	99.58 -116 - 315 (109.86) [0.36] {0.98}	587.61 (344.20)	60.55 -122 - 243 (92.23) [0.51] {0.99}
Urgent Care	25.51 (108.94)	-5.73 -13 - 2 (3.82) [0.14] {0.79}	-9.03 -21 - 3 (5.95) [0.13] {0.78}	25.02 (104.16)	-4.95 -13 - 3 (4.17) [0.24] {0.92}	-7.69 -20 - 5 (6.41) [0.23] {0.92}	26.91 (20.25)	-5.73 -15 - 4 (4.82) [0.24] {0.92}
Other	116.92 (1336.48)	-42.22 -105 - 20 (31.59) [0.18] {0.88}	-66.48 -162 - 29 (48.78) [0.17] {0.86}	115.34 (1399.24)	-76.28 -111 - -41 (17.72) [0.00] {0.01}	-118.52 -172 - -65 (27.19) [0.00] {0.01}	169.61 (285.46)	-76.77 -204 - 50 (64.31) [0.23] {0.92}
By Site of Care:								
Any Physician Visit (%)	71.81 (45.00)	0.31 -3 - 4 (1.82) [0.87] {1.00}	0.48 -5 - 6 (2.85) [0.87] {1.00}	75.08 (43.26)	0.76 -3 - 4 (1.85) [0.68] {0.99}	1.18 -4 - 7 (2.87) [0.68] {0.99}	71.75 (7.90)	0.57 -3 - 4 (1.87) [0.76] {1.00}
Number of Physician Visits	3.22 (4.12)	0.11 -0.16 - 0.38 (0.14) [0.44] {0.97}	0.17 -0.26 - 0.59 (0.22) [0.44] {0.97}	3.27 (4.05)	0.05 -0.22 - 0.31 (0.13) [0.73] {0.99}	0.07 -0.34 - 0.48 (0.21) [0.73] {0.99}	3.58 (0.92)	0.01 -0.35 - 0.38 (0.18) [0.95] {1.00}

	Employee-level			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)
Any Hospitalization (%)	6.66 (24.94)	-0.97 -3 - 1 (0.88) [0.27] {0.94}	-1.53 -4 - 1 (1.36) [0.26] {0.94}	6.60 (24.83)	-0.70 -3 - 1 (1.08) [0.52] {0.99}	-1.08 -4 - 2 (1.65) [0.51] {0.99}	6.67 (3.74)	-0.66 -3 - 1 (0.99) [0.51] {0.99}
Number of Hospitalizations	0.07 (0.32)	-0.02 -0.03 – 0.00 (0.01) [0.08] {0.67}	-0.02 -0.05 – 0.00 (0.01) [0.07] {0.64}	0.06 (0.30)	-0.01 -0.03 – 0.00 (0.01) [0.10] {0.68}	-0.02 -0.05 – 0.00 (0.01) [0.09] {0.66}	0.09 (0.06)	-0.02 -0.05 - 0.01 (0.01) [0.16] {0.85}
Any ER Visit (%)	21.52 (41.10)	-0.38 -3 - 3 (1.58) [0.81] {1.00}	-0.60 -5 - 4 (2.47) [0.81] {1.00}	22.17 (41.54)	0.19 -3 - 4 (1.81) [0.92] {0.99}	0.30 -5 - 6 (2.80) [0.92] {0.99}	21.49 (7.50)	-0.06 -3 - 3 (1.71) [0.97] {1.00}
Number of ER Visits	0.26 (0.67)	0.02 -0.04 - 0.09 (0.03) [0.47] {0.97}	0.04 -0.06 - 0.13 (0.05) [0.46] {0.97}	0.23 (0.56)	0.02 -0.03 - 0.07 (0.02) [0.43] {0.99}	0.03 -0.04 - 0.10 (0.04) [0.42] {0.98}	0.28 (0.12)	0.02 -0.05 - 0.08 (0.03) [0.58] {0.99}
Any Urgent Care Visit (%)	13.13 (33.77)	-1.91 -5 - 1 (1.67) [0.26] {0.93}	-3.01 -8 - 2 (2.61) [0.25] {0.94}	13.68 (34.37)	-1.74 -6 - 2 (2.00) [0.38] {0.98}	-2.71 -9 - 3 (3.08) [0.38] {0.98}	13.36 (7.55)	-2.27 -6 - 2 (2.04) [0.27]
Number of Urgent Care Visits	0.14 (0.47)	-0.02 -0.06 - 0.02 (0.02) [0.40] {0.97}	-0.03 -0.1 - 0.04 (0.03) [0.39] {0.97}	0.14 (0.44)	-0.01 -0.06 - 0.04 (0.03) [0.62] {0.99}	-0.02 -0.10 - 0.06 (0.04) [0.61] {0.99}	0.15 (0.10)	-0.03 -0.08 - 0.03 (0.03) [0.32] {0.95}
Any Preventive Care Visit (%)	35.98 (48.00)	-0.64 -7 - 6 (3.35) [0.85] {1.00}	-1.00 -11 - 9 (5.23) [0.85] {1.00}	38.81 (48.74)	-0.41 -8 - 7 (3.70) [0.91] {0.99}	-0.64 -12 - 11 (5.72) [0.91] {0.99}	35.96 (10.78)	-1.70 -8 - 4 (3.13) [0.59] {0.99}
Number of Preventive Care Visits	0.36 (0.57)	0.01 -0.07 - 0.08 (0.04) [0.85] {1.00}	0.01 -0.1 - 0.12 (0.06) [0.85] {1.00}	0.36 (0.53)	0.01 -0.07 - 0.09 (0.04) [0.77] {0.99}	0.02 -0.10 - 0.14 (0.06) [0.77] {0.99}	0.38 (0.13)	-0.01 -0.08 - 0.06 (0.04) [0.86] {1.00}
N	6626	7631	7631	5222	6016	6016	140	160

Notes: This table reports the coefficient on TREATMENT from estimating equation (1) by OLS (column 2), and the coefficient on PARTICIPATION from estimating equation (2) by IV (column 3). Standard errors are listed in parentheses with p-values in brackets and family-wise p-values in curly braces. Column 1 reports the mean of each medical spending and utilization outcome in the control group for each sample (with standard deviation in parentheses). All regressions included demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, full-time status, paid hourly status, and job category) and clustered standard errors by the worksite (for employee-level regressions). The employee-level control means and regressions were weighted by the combination of a weight for exposure to the wellness program and a weight that balances treatment and control samples on demographics. We multiplied individual-adjusted worksite-level outcomes by 2000 to allow the data

to approximate the magnitude of a full-time equivalent employee (2000 hours). Worksite-level control means and regressions were unweighted.

eTable 9. Mean Values and Effect of Program on Prescription Drug Spending and Utilization

	All Employees			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)
Prescription Drug Spending								
Total Spending	1214.51 (7423.72)	179.40 -245 - 603 (214.64) [0.40] {0.99}	282.50 -377 - 942 (336.72) [0.40] {0.99}	1142.41 (5329.42)	230.62 -200 - 661 (217.90) [0.29] {0.97}	358.31 -308 - 1025 (340.06) [0.29] {0.97}	1279.26 (1006.67)	187.70 -271 - 647 (232.38) [0.42] {0.99}
Out-of-pocket Spending	93.54 (169.77)	7.05 -5 - 19 (6.20) [0.26] {0.93}	11.09 -8 - 30 (9.71) [0.25] {0.93}	98.00 (171.97)	7.72 -7 - 22 (7.21) [0.29] {0.97}	11.99 -10 - 34 (11.19) [0.28] {0.97}	103.44 (31.80)	8.44 -7 - 24 (7.90) [0.29] {0.96}
Prescription Drug Utilization								
Any Medications (%)	58.55 (49.27)	2.09 -1 - 6 (1.74) [0.23] {0.93}	3.29 -2 - 9 (2.71) [0.22] {0.93}	61.51 (48.66)	3.12 -1 - 7 (1.93) [0.11] {0.75}	4.85 -1 - 11 (3.00) [0.11] {0.75}	58.60 (8.57)	1.52 -2 - 5 (1.89) [0.42] {0.75}
Number of Distinct Medications	4.01 (4.74)	0.25 0 - 1 (0.16) [0.12] {0.80}	0.40 0 - 1 (0.26) [0.12] {0.80}	4.28 (4.80)	0.33 0 - 1 (0.18) [0.06] {0.63}	0.51 0 - 1 (0.28) [0.06] {0.63}	4.01 (0.90)	0.25 0 - 1 (0.19) [0.20] {0.92}
Number of Medication Months	11.04 (19.71)	0.60 -1 - 2 (0.73) [0.41] {0.99}	0.95 -1 - 3 (1.17) [0.42] {0.99}	11.59 (20.10)	0.80 -1 - 2 (0.85) [0.35] {0.97}	1.24 -1 - 4 (1.34) [0.35] {0.97}	12.44 (4.46)	0.70 -1 - 2 (0.85) [0.41] {0.99}
By Clinical Category:								
Any Asthma Medications (%)	11.79 (32.25)	2.05 -1 - 5 (1.41) [0.15] {0.85}	3.22 -1 - 8 (2.22) [0.15] {0.85}	12.59 (33.18)	2.65 0 - 6 (1.53) [0.09] {0.71}	4.12 -1 - 9 (2.38) [0.08] {0.71}	11.75 (5.23)	2.66 -1 - 6 (1.61) [0.10] {0.78}
Number of Asthma Medication Months	0.51 (2.49)	-0.01 -0.15 - 0.13 (0.07) [0.87] {1.00}	-0.02 -0.23 - 0.19 (0.11) [0.86] {1.00}	0.52 (2.49)	0.02 -0.13 - 0.17 (0.08) [0.80] {1.00}	0.03 -0.2 - 0.27 (0.12) [0.80] {1.00}	0.56 (0.44)	0.02 -0.15 - 0.18 (0.08) [0.83] {1.00}
Any Cardiovascular Medications (%)	22.31 (41.64)	0.40 -2 - 3 (1.42) [0.78] {1.00}	0.63 -4 - 5 (2.22) [0.78] {1.00}	23.87 (42.63)	0.87 -3 - 4 (1.73) [0.61] {1.00}	1.35 -4 - 7 (2.69) [0.61] {1.00}	22.09 (7.76)	0.84 -2 - 4 (1.56) [0.59] {1.00}
Number of Cardiovascular	2.55 (6.52)	-0.01 -0.47 - 0.46	-0.01 -0.73 - 0.71	2.72 (6.68)	0.09 -0.48 - 0.65	0.13 -0.74 - 1.01	2.84 (1.28)	0.08 -0.53 - 0.7

	All Employees			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)
Medication Months		(0.24) [0.98] {1.00}	(0.37) [0.98] {1.00}		(0.29) [0.76] {1.00}	(0.45) [0.76] {1.00}		(0.31) [0.79] {1.00}
Any Diabetes Medications (%)	7.07 (25.64)	0.56 -1 - 2 (0.89) [0.53] {1.00}	0.89 -2 - 4 (1.40) [0.53] {1.00}	7.49 (26.33)	0.82 -1 - 3 (0.98) [0.40] {0.98}	1.28 -2 - 4 (1.53) [0.40] {0.98}	6.93 (3.99)	0.39 -2 - 2 (0.98) [0.69] {1.00}
Number of Diabetes Medication Months	0.95 (4.54)	0.06 -0.24 - 0.35 (0.15) [0.71] {1.00}	0.09 -0.38 - 0.55 (0.24) [0.71] {1.00}	1.02 (4.72)	0.06 -0.26 - 0.38 (0.16) [0.71] {1.00}	0.09 -0.39 - 0.58 (0.25) [0.71] {1.00}	1.04 (0.81)	0.02 -0.33 - 0.37 (0.18) [0.91] {1.00}
Any Hyperlipidemia Medications (%)	13.95 (34.65)	-0.27 -2 - 2 (1.09) [0.80] {1.00}	-0.43 -4 - 3 (1.71) [0.80] {1.00}	15.05 (35.76)	-0.54 -3 - 2 (1.31) [0.68] {1.00}	-0.85 -5 - 3 (2.01) [0.67] {1.00}	13.72 (6.20)	0.22 -2 - 2 (1.12) [0.84] {1.00}
Number of Hyperlipidemia Medication Months	1.14 (3.49)	-0.04 -0.23 - 0.15 (0.10) [0.70] {1.00}	-0.06 -0.35 - 0.23 (0.15) [0.69] {1.00}	1.21 (3.57)	-0.06 -0.28 - 0.16 (0.11) [0.59] {1.00}	-0.09 -0.43 - 0.24 (0.17) [0.59] {1.00}	1.26 (0.70)	-0.01 -0.26 - 0.24 (0.12) [0.94] {1.00}
Any Mental Health Medications (%)	17.44 (37.94)	1.20 -2 - 5 (1.72) [0.49] {1.00}	1.89 -3 - 7 (2.71) [0.48] {1.00}	18.12 (38.52)	1.97 -2 - 6 (2.07) [0.34] {0.97}	3.06 -3 - 9 (3.23) [0.34] {0.97}	17.83 (7.22)	1.62 -2 - 5 (1.90) [0.40] {0.99}
Number of Mental Health Medication Months	1.65 (5.30)	0.13 -0.22 - 0.48 (0.18) [0.47] {1.00}	0.20 -0.35 - 0.75 (0.28) [0.47] {1.00}	1.70 (5.37)	0.24 -0.20 - 0.67 (0.22) [0.28] {0.97}	0.37 -0.31 - 1.04 (0.34) [0.29] {0.97}	1.92 (1.19)	0.10 -0.37 - 0.58 (0.24) [0.67] {1.00}
Any Pain Medications (%)	17.61 (38.09)	2.43 0 - 5 (1.42) [0.09] {0.71}	3.82 0 - 8 (2.16) [0.08] {0.68}	18.67 (38.97)	3.14 0 - 6 (1.66) [0.06] {0.63}	4.88 0 - 10 (2.48) [0.05] {0.60}	17.60 (7.20)	0.85 -2 - 4 (1.36) [0.53] {1.00}
Number of Pain Medication Months	0.75 (2.74)	0.02 -0.13 - 0.18 (0.08) [0.76] {1.00}	0.04 -0.20 - 0.28 (0.12) [0.75] {1.00}	0.76 (2.78)	0.01 -0.18 - 0.20 (0.10) [0.90] {1.00}	0.02 -0.27 - 0.31 (0.15) [0.90] {1.00}	0.87 (0.60)	-0.10 -0.29 - 0.08 (0.09) [0.26] {0.96}
Any Antibiotics Medications (%)	12.84 (33.45)	-0.18 -3 - 2 (1.32) [0.89] {1.00}	-0.28 -4 - 4 (2.07) [0.89] {1.00}	14.14 (34.85)	-0.10 -3 - 3 (1.52) [0.95] {1.00}	-0.16 -5 - 4 (2.35) [0.95] {1.00}	12.74 (5.55)	-0.77 -4 - 2 (1.54) [0.62] {1.00}
Number of Antibiotics Medication Months	0.39 (1.60)	0.03 -0.07 - 0.13 (0.05) [0.57] {1.00}	0.05 -0.12 - 0.21 (0.08) [0.57] {1.00}	0.40 (1.64)	-0.01 -0.10 - 0.09 (0.05) [0.89] {1.00}	-0.01 -0.15 - 0.13 (0.07) [0.89] {1.00}	0.43 (0.27)	-0.04 -0.13 - 0.05 (0.05) [0.35] {0.99}
Any Other Medications	34.26 (47.46)	3.42 0 - 7	5.39 1 - 10	36.33 (48.10)	3.37 0 - 7	5.24 0 - 10	34.30 (8.03)	2.58 -1 - 6

	All Employees			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat	Local Average Treatment Effect	Mean Value in Control Group	Intent-to-treat
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)
(%)		(1.56) [0.03] {0.43}	(2.46) [0.03] {0.45}		(1.70) [0.05] {0.58}	(2.68) [0.05] {0.60}		(1.74) [0.14] {0.85}
Number of Other Medication Months	3.10 (7.09)	0.42 0 - 1 (0.34) [0.22] {0.93}	0.66 0 - 2 (0.54) [0.22] {0.93}	3.25 (7.24)	0.45 0 - 1 (0.39) [0.24] {0.95}	0.71 0 - 2 (0.61) [0.25] {0.95}	3.50 (1.31)	0.63 0 - 1 (0.34) [0.07] {0.65}
N	6626	7631	7631	5222	6016	6016	140	160

Notes: This table reports the coefficient on TREATMENT from estimating equation (1) by OLS (column 2), and the coefficient on PARTICIPATION from estimating equation (2) by IV (column 3). Standard errors are listed in parentheses with p-values in brackets and family-wise p-values in curly braces. Column 1 reports the mean of each prescription drug spending and utilization outcome in the control group for each sample (with standard deviation in parentheses). All regressions included demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, full-time status, paid hourly status, and job category) and clustered standard errors by the worksite (for employee-level regressions). Employee-level control means and regressions are weighted by the combination of a weight for exposure to the wellness program and a weight that balances treatment and control samples on demographics. We multiplied individual-adjusted worksite-level outcomes by 2000 to allow the data to approximate the magnitude of a full-time equivalent employee (2000 hours). Worksite-level control means and regressions were unweighted.

eTable 10. Mean Values and Effect of Program on Employment Outcomes

	All Employees			Stably Employed Subsample			Worksite-level	
	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)	Local Average Treatment Effect (3)	Mean Value in Control Group (1)	Intent-to-treat (2)
Absenteeism (% of scheduled hours missed)	2.63 (1.64)	-0.14 -0.30 - 0.02 (0.08) [0.09] {0.21}	-0.25 -0.52 - 0.03 (0.14) [0.08] {0.20}	2.86 (1.62)	-0.13 -0.29 - 0.04 (0.08) [0.14] {0.30}	-0.20 -0.47 - 0.06 (0.13) [0.13] {0.27}	2.57 (0.36)	-0.13 -0.30 - 0.05 (0.09) [0.16] {0.33}
Performance Review (% with a score better than 3 out of 5)	60.53 (48.88)	-0.46 -8.32 - 7.40 (3.98) [0.91] {0.92}	-0.78 -14.00 - 12.44 (6.74) [0.91] {0.92}	66.53 (47.19)	-0.06 -7.63 - 7.50 (3.83) [0.99] {0.99}	-0.10 -12.20 - 11.99 (6.17) [0.99] {0.99}	59.05 (13.72)	-1.94 -8.81 - 4.93 (3.48) [0.58] {0.60}
Tenure (days during treatment; for site-level: % of treatment period)[†]	308.81 (212.56)	-5.55 -18.81 - 7.70 (6.71) [0.41] {0.45}	-15.81 -53.14 - 21.51 (19.05) [0.41] {0.45}	460.77 (151.16)	-0.11 -0.30 - 0.09 (0.10) [0.28] {0.99}	-0.21 -0.57 - 0.16 (0.19) [0.27] {0.99}	85.40 (3.62)	-1.52 -2.91 - -0.13 (0.70) [0.03] {0.13}
N	21079 - 28940	24054 - 32974		13366 - 13452	15246 - 15344		140	160

Notes: This table reports the coefficient on TREATMENT from estimating equation (1) by OLS (column 2), and the coefficient on PARTICIPATION from estimating equation (2) by IV (column 3). Standard errors are listed in parentheses with p-values in brackets and family-wise p-values in curly braces. Column 1 reports the mean of each prescription drug spending and utilization outcome in the control group for each sample (with standard deviation in parentheses). All regressions included demographic and employment covariates (age, sex, age-sex interactions, race/ethnicity, Cigna coverage status, full-time status, paid hourly status, and job category) and clustered standard errors by worksite (for employee-level regressions). Control means and regressions for tenure were weighted by a weight that balances treatment and control groups on demographics. Control means and regressions for absenteeism and performance review were weighted by the combination of this weight and a weight for exposure to the wellness program. Multiple inference adjustment was performed for absenteeism and performance review in a combined manner consisted with other results using the family-wise p-values; it was separately performed for tenure given the difference in weights for the tenure regression. Worksite-level control means and regressions were unweighted.

[†]For worksite level results on tenure, the outcome was defined as the percent of total days at the worksite during the study period that was worked by employees at the worksite.

eTable 11. Heterogeneity Analyses of Key Pre-Specified Outcomes

A. Heterogeneity by Sex: Females vs. Males

	Mean Value in Control Group		Intent-to-treat			Local Average Treatment Effect		
	Female	Male	Female	Male	P value of difference	Female	Male	P value of difference
Total Spending	4727.90 (13044.33)	3329.91 (15876.95)	-249.89 (645.25)	-518.03 (460.54)	0.70	-312.27 (798.04)	-1035.11 (899.19)	0.50
Total Rx Spending	1238.66 (6286.85)	1195.09 (8225.50)	490.12 (418.08)	-58.69 (280.68)	0.32	610.57 (515.07)	-117.54 (560.18)	0.39
Absenteeism (%)	2.86 (1.58)	2.43 (1.66)	-0.18 (0.10)	-0.11 (0.09)	0.44	-0.25 (0.14)	-0.24 (0.19)	0.96
Systolic BP	120.59 (17.23)	128.99 (15.18)	1.71 (1.13)	-1.73 (1.56)	0.07	2.05 (1.34)	-2.43 (2.16)	0.07
BMI	30.31 (7.33)	28.93 (6.70)	0.27 (0.45)	-0.13 (0.53)	0.56	0.33 (0.54)	-0.19 (0.75)	0.58
Annual exam (%)	72.22 (44.83)	57.14 (49.55)	-3.72 (3.07)	2.26 (4.88)	0.29	-4.48 (3.67)	3.13 (6.74)	0.30
SF-8 mental	50.16 (9.77)	52.44 (7.96)	-0.21 (0.55)	-0.56 (0.85)	0.74	-0.25 (0.65)	-0.77 (1.15)	0.71
SF-8 physical	49.93 (8.01)	51.87 (7.18)	-0.03 (0.51)	-0.35 (0.55)	0.68	-0.03 (0.61)	-0.49 (0.75)	0.66
Regular exercise (%)	56.89 (49.57)	68.12 (46.66)	9.96 (3.15)	6.44 (3.58)	0.49	12.04 (3.71)	8.94 (4.78)	0.63
Actively managing weight (%)	59.85 (49.07)	47.90 (50.02)	12.65 (4.52)	15.46 (3.70)	0.60	15.24 (5.38)	21.06 (4.97)	0.37
Sweetened drinks per day	1.67 (1.72)	2.07 (2.00)	0.08 (0.15)	0.14 (0.17)	0.79	0.10 (0.17)	0.20 (0.23)	0.73
Smoker (%)	25.03 (43.36)	24.14 (42.84)	-8.22 (3.72)	-5.31 (3.66)	0.51	-9.89 (4.46)	-7.44 (5.12)	0.67
Alcoholic drinks per week	3.41 (6.07)	6.20 (8.58)	-0.24 (0.38)	-1.09 (0.57)	0.26	-0.28 (0.46)	-1.52 (0.79)	0.22

Notes: This table reports heterogeneity in treatment effects using intent-to-treat and local average treatment effect analyses by sex, testing for differences in the coefficient of interest between males and females. For control group means, standard deviations are shown in parentheses. For treatment effect estimates, standard errors are shown in parentheses. P values of the difference in effects between males and females are reported in a separate column for both the intent-to-treat and local average treatment effect estimates. All regressions included the same covariates and weights as their primary specifications and similarly clustered standard errors by worksite. The "Actively managing weight" outcome was added to heterogeneity analyses after our pre-specified analysis plan was finalized.

B. Heterogeneity by Age: Below 40 vs. 40 and Above

	Mean Value in Control Group		Intent-to-treat			Local Average Treatment Effect		
	Below 40	40 and Above	Below 40	40 and Above	P value of difference	Below 40	40 and Above	P value of difference
Total Spending	2383.42 (13018.62)	4984.47 (15619.19)	-22.94 (496.60)	-688.30 (612.64)	0.40	-35.92 (780.92)	-1080.99 (941.11)	0.39
Total Rx Spending	838.98 (8557.68)	1461.29 (6562.31)	-9.88 (289.08)	302.91 (285.61)	0.42	-15.76 (455.12)	475.71 (449.26)	0.42
Absenteeism (%)	2.48 (1.47)	2.80 (1.79)	-0.15 (0.09)	-0.13 (0.10)	0.77	-0.27 (0.15)	-0.22 (0.16)	0.69
Blood Pressure, Systolic	119.34 (14.71)	128.82 (17.46)	-0.73 (1.38)	1.07 (1.20)	0.30	-0.94 (1.77)	1.35 (1.51)	0.29
BMI	29.12 (7.17)	30.23 (6.99)	0.40 (0.48)	-0.17 (0.51)	0.42	0.52 (0.62)	-0.21 (0.64)	0.41
Annual exam (%)	52.27 (50.00)	77.41 (41.87)	5.74 (4.63)	-7.17 (3.40)	0.02	7.43 (5.97)	-9.00 (4.22)	0.02
SF-8 mental	49.65 (9.91)	52.56 (8.01)	-0.66 (0.84)	-0.07 (0.48)	0.56	-0.85 (1.07)	-0.10 (0.59)	0.55
SF-8 physical	51.96 (7.48)	49.71 (7.77)	-0.69 (0.50)	0.32 (0.53)	0.20	-0.89 (0.64)	0.39 (0.66)	0.19
Regular exercise (%)	67.61 (46.84)	56.52 (49.63)	-0.15 (3.40)	15.74 (4.16)	0.01	-0.19 (4.36)	19.76 (5.05)	0.01
Actively managing weight (%)	55.28 (49.77)	54.18 (49.89)	10.69 (4.37)	16.07 (4.17)	0.32	13.70 (5.56)	20.14 (5.25)	0.35
Sweetened drinks per day	2.14 (1.97)	1.58 (1.71)	0.25 (0.19)	-0.02 (0.11)	0.23	0.33 (0.25)	-0.02 (0.14)	0.22
Smoker (%)	24.54 (43.07)	24.72 (43.18)	-7.45 (3.34)	-6.37 (3.95)	0.80	-9.70 (4.32)	-8.03 (4.93)	0.76
Alcoholic drinks per week	4.79 (7.34)	4.51 (7.49)	-0.49 (0.54)	-0.65 (0.45)	0.84	-0.63 (0.70)	-0.82 (0.56)	0.86

Notes: This table reports heterogeneity in treatment effects using intent-to-treat and local average treatment effect analyses by age, testing for differences in the coefficient of interest between individuals younger than 40 years old and individuals 40 years or older. For control group means, standard deviations are shown in parentheses. For treatment effect estimates, standard errors are shown in parentheses. P values of the difference in effects between the age subgroups are reported in a separate column for both the intent-to-treat and local average treatment effect estimates. All regressions included the same covariates and weights as their primary specifications and similarly clustered standard errors by worksite. Consistent with Table A above, the "Actively managing weight" outcome was added to heterogeneity analyses after our pre-specified analysis plan was finalized.

eTable 12. Alternative Definitions of Participation

	Intent-to-Treat (All Employees)	Local Average Treatment Effect using alternative definitions of participation		
		Any Module	Three or More Modules	Number of Modules
	(1)	2SLS (2)	2SLS (3)	2SLS (4)
Total medical spending	-425.57 (425.73) [0.32]	-670.13 (655.05) [0.31]	-867.12 (843.42) [0.30]	-143.37 (139.40) [0.30]
Total Rx spending	179.40 (214.64) [0.40]	282.50 (336.72) [0.40]	365.54 (439.55) [0.41]	60.44 (72.44) [0.40]
Absenteeism (%)	-0.14 (0.08) [0.09]	-0.25 (0.14) [0.08]	-0.35 (0.20) [0.08]	-0.06 (0.03) [0.08]
Blood Pressure, Systolic	0.24 (0.96) [0.81]	0.30 (1.22) [0.80]	0.37 (1.50) [0.81]	0.06 (0.25) [0.81]
BMI	0.09 (0.34) [0.79]	0.12 (0.44) [0.78]	0.15 (0.54) [0.78]	0.02 (0.09) [0.78]
Annual exam (%)	-1.26 (2.87) [0.66]	-1.61 (3.62) [0.66]	-1.98 (4.46) [0.66]	-0.33 (0.74) [0.66]
SF-8 mental	-0.15 (0.35) [0.66]	-0.19 (0.44) [0.66]	-0.24 (0.54) [0.66]	-0.04 (0.09) [0.66]
SF-8 physical	-0.35 (0.45) [0.44]	-0.44 (0.57) [0.43]	-0.55 (0.70) [0.43]	-0.09 (0.11) [0.43]
Regular exercise (%)	8.33 (2.25) [0.00]	10.64 (2.75) [0.00]	13.10 (3.46) [0.00]	2.17 (0.58) [0.00]
Sweetened drinks per day	0.10 (0.11) [0.34]	0.13 (0.14) [0.33]	0.17 (0.17) [0.33]	0.03 (0.03) [0.33]
Smoker (%)	-6.87 (3.01) [0.03]	-8.78 (3.84) [0.02]	-10.78 (4.81) [0.02]	-1.78 (0.79) [0.02]
Alcoholic drinks per week	-0.57 (0.28) [0.04]	-0.73 (0.36) [0.04]	-0.90 (0.44) [0.04]	-0.15 (0.07) [0.04]

Notes: This table replicates the intent-to-treat estimates from the full sample (column 1) and compares them to the local average treatment effect estimates (2SLS) using our baseline definition of participation (column 2) alongside our alternate definitions of participation (columns 3-4). Of note, participation in 3 or more modules (column 3) can be thought of as a more intensive degree of participation compared to participation in any module (column 2), which helps lend a dose-response type of interpretation when columns 2 and 3 are considered together. Standard errors are shown in parentheses and p-values are shown in brackets. All regressions included the same covariates and weights as their primary specifications and similarly clustered standard errors by worksite.

eTable 13. Exposure Weights Only

	Intent-to-treat (1)	Local Average Treatment Effect (2)
Total medical spending	-479.34 (413.93) [0.25]	-765.41 (645.20) [0.24]
Total Rx spending	128.40 (215.26) [0.55]	205.03 (342.79) [0.55]
Absenteeism	-0.11 (0.08) [0.18]	-0.19 (0.14) [0.17]
Blood Pressure, Systolic	0.31 (0.92) [0.74]	0.40 (1.18) [0.74]
BMI	0.12 (0.35) [0.73]	0.15 (0.44) [0.73]
Annual exam	-1.34 (2.95) [0.65]	-1.73 (3.74) [0.64]
SF-8 physical	-0.16 (0.34) [0.65]	-0.20 (0.43) [0.64]
SF-8 mental	-0.31 (0.42) [0.46]	-0.40 (0.54) [0.46]
Regular exercise	7.57 (2.27) [0.00]	9.74 (2.81) [0.00]
Number of sweetened drinks per day	0.11 (0.10) [0.28]	0.15 (0.13) [0.28]
Smoking	-6.53 (2.99) [0.03]	-8.39 (3.84) [0.03]
Alcohol drinks per week	-0.58 (0.28) [0.04]	-0.75 (0.36) [0.04]

Note: This table reports results for a key set of outcomes when using only exposure weights rather than the combination of exposure weights and balance weights. In all other respects, the regression models are identical to the base specification in prior tables. All regressions included the same covariates as their primary specifications and clustered standard errors by worksite. Standard errors are reported in parentheses and p-values in brackets.

eTable 14. Logit Model Estimates for Binary Outcome Variables

	All Employees	Stably Employed Sub-Sample
	Reduced Form	Reduced Form
Self-Reported Health and Behaviors		
Annual exam (%)	0.23 (2.98) [0.94]	-1.84 (3.42) [0.59]
Flu shot (%)	-3.38 (3.39) [0.32]	-2.95 (3.22) [0.36]
Percent of other recommended tests received	4.12 (1.70) [0.02]	3.55 (1.76) [0.04]
PHQ-2 score of 3 or above (%)	-1.10 (1.02) [0.28]	-0.71 (1.26) [0.58]
Unmanaged stress (%)	-1.61 (2.23) [0.47]	-3.00 (2.90) [0.30]
Stress at work (%)	0.26 (2.40) [0.92]	3.31 (2.51) [0.19]
Good quality, adequate amount of sleep (%)	-0.63 (2.17) [0.77]	-2.92 (2.47) [0.24]
Regular exercise (%)	8.82 (2.15) [0.00]	8.83 (2.50) [0.00]
3 or more days of moderate exercise per week (%)	2.69 (2.18) [0.22]	3.71 (2.72) [0.17]
Read the Nutrition Facts panel (%)	2.81 (2.45) [0.25]	8.04 (3.22) [0.01]
Consume at least 2 cups of fruit and 2.5 cups of vegetables per day (%)	4.44 (2.21) [0.04]	5.07 (2.78) [0.07]
Choose whole grain foods and reduced fat foods more often than the regular variety (%)	1.96 (1.97) [0.32]	1.57 (2.67) [0.56]
Considering losing weight in the next 6 months (%)	10.86 (3.20) [0.00]	9.68 (3.00) [0.00]
Actively managing weight (%)	13.09 (3.18)	14.11 (3.50)

	All Employees Reduced Form	Stably Employed Sub-Sample Reduced Form
	[0.00]	[0.00]
Smoker (%)	-6.19 (2.77) [0.03]	-8.85 (3.25) [0.01]
Any doctor visit in last 12 months (%)	0.20 (2.21) [0.93]	-0.70 (2.78) [0.80]
Any ER visit in last 12 months (%)	-2.93 (2.18) [0.18]	-3.66 (2.58) [0.16]
Ever hospital patient in the last 12 months (%)	-1.56 (1.93) [0.42]	-3.27 (2.40) [0.17]
Any prescriptions in last 12 months (%)	-2.28 (2.33) [0.33]	-0.98 (2.67) [0.71]
Clinical Measures of Health (Biometrics)		
High total cholesterol (total cholesterol ≥200 mg/dl)	0.64 (3.86) [0.87]	-0.51 (4.89) [0.92]
Low HDL cholesterol (HDL cholesterol <40 mg/dl)	-1.06 (2.18) [0.63]	0.01 (2.71) [1.00]
Hypertension (systolic BP ≥140 or diastolic BP ≥90 mm Hg)	2.73 (2.18) [0.21]	3.57 (2.93) [0.22]
Obesity (BMI ≥30)	1.03 (2.25) [0.65]	1.39 (2.50) [0.58]
Medical Spending and Utilization		
Any Physician Visit (%)	-0.34 (1.76) [0.85]	0.88 (1.95) [0.65]
Any Hospitalization (%)	-0.98 (0.85) [0.25]	-0.68 (1.06) [0.52]
Any ER Visit (%)	-0.80 (1.40) [0.57]	0.21 (1.81) [0.91]
Any Urgent Care Visit (%)	-1.76 (1.69) [0.30]	-1.79 (2.11) [0.40]
Any Preventive Care Visit (%)	-1.12 (3.11) [0.72]	-0.47 (4.00) [0.91]

	All Employees Reduced Form	Stably Employed Sub-Sample Reduced Form
Prescription Drugs Spending and Utilization		
Any Medications (%)	0.56 (1.62) [0.73]	3.55 (2.20) [0.11]
Any Asthma Medications (%)	1.19 (1.05) [0.26]	2.35 (1.30) [0.07]
Any Cardiovascular Medications (%)	0.13 (1.40) [0.93]	0.90 (1.63) [0.58]
Any Diabetes Medications (%)	0.07 (0.61) [0.90]	0.67 (0.72) [0.35]
Any Hyperlipidemia Medications (%)	-0.05 (0.64) [0.93]	-0.35 (0.95) [0.71]
Any Mental Health Medications (%)	1.02 (1.45) [0.48]	1.82 (1.84) [0.33]
Any Pain Medications (%)	0.77 (1.24) [0.53]	2.85 (1.43) [0.05]
Any Antibiotics Medications (%)	-0.62 (1.13) [0.58]	-0.16 (1.46) [0.91]
Any Other Medications (%)	2.00 (1.55) [0.20]	3.59 (1.83) [0.05]
Employment		
Performance Review (% with a score better than 3 out of 5)	-0.95 (4.37) [0.83]	0.11 (3.96) [0.98]

Notes: To test the robustness of our results from the linear model, we estimated logit models for binary outcome variables. This table reports the coefficient of interest in a reduced form logit model for each binary outcome we evaluated, transformed to be interpretable as a marginal effect. Standard errors are reported in parentheses and p-values in brackets. All regressions included the same covariates and weights as their primary specifications and similarly clustered standard errors by worksite. Results are reported for all employees and for the stably employed subsample.

eTable 15. Analysis of Potential Selection: Participants vs. Non-Participants in Program Modules

A. All Individuals

	(1) Participation in any module			(2) Participation in at least 3 modules		
	Non-participants (n=2,616)	**Participants (n=1,421)**	**P value**	**Non-participants (n=3,172)**	**Participants (n=865)**	**P value**
Age (yrs)	38.91	40.10	0.23	38.32	41.47	0.01
Female (%)	31.37	57.43	<0.001	36.04	61.62	<0.001
Race (%)			<0.001			0.04
White	60.21	53.39		58.75	52.64	
Black	19.69	19.89		19.35	20.48	
Hispanic	13.22	21.32		15.75	21.02	
Other race	6.88	5.41		6.15	5.87	
Employment (%)			<0.001			<0.001
Full-time salary	10.19	14.08		10.66	15.05	
Full-time hourly	50.13	45.22		47.96	46.34	
Part-time hourly	39.68	40.71		41.38	38.62	
Worker Type (%)			<0.001			<0.001
Sales worker	29.92	42.62		33.36	42.94	
Nonsales worker	59.51	37.68		54.49	35.81	
Other worker	10.57	19.70		12.15	21.25	

B. Individuals with Cigna coverage

	(1) Participation in any module			(2) Participation in at least 3 modules		
Healthcare Spending in 2014	**Non-participants (n=334)**	**Participants (n=470)**	**P value**	**Non-participants (n= 447)**	**Participants (n=357)**	**P value**
Medical Spending ($)	3356.97	4961.90	0.33	3059.15	5722.40	0.18
Any Spending >$0 (%)	72.82	72.52	0.96	69.32	75.95	0.09
Drug Spending ($)	897.01	1054.42	0.51	833.90	1163.09	0.15
Any Spending >$0 (%)	67.14	68.61	0.79	64.49	71.71	0.06

Note: These tables report the results of a retrospective selection analysis that compares the demographic and employment characteristics, as well as pre-intervention (2014) medical and drug spending, of participants and non-participants within the treatment group. Participants were employees randomized into treatment worksites who participated in the wellness program. Non-participants were employees randomized into treatment worksites who did not participate in the program. We used 2 of our definitions of participation, which captured a measure of the intensity of participation. The first set of columns (1) defines participation as completion of at least 1 module of the wellness program; the second set of columns (2) defines participation as completion of 3 or more modules. Panel A includes all individuals. Panel B includes the subset of individuals who have employer-sponsored health insurance coverage through Cigna, for whom we have claims data.

eTable 16. Analysis of Potential Selection: Participants vs. Non-Participants in Survey/Biometrics

Variable	Treatment Worksites (20)				Primary Control Worksites (20)				Between-Group Difference	
	Participants (P)	Non-participants (NP)	Difference (P- NP)	p-value	Participants (P)	Non-participants (NP)	Difference (P- NP)	p-value	ΔT–ΔC	p-value
Age (yrs)	40.05	34.71	5.34	<0.001	38.38	33.61	4.77	<0.001	0.57	0.466
Female (%)	0.60	0.44	0.16	<0.001	0.58	0.44	0.13	<0.001	0.028	0.263
White race (%)	0.68	0.63	0.05	0.008	0.58	0.52	0.06	0.001	-0.011	0.647
Full-time employed (%)	0.62	0.54	0.09	<0.001	0.64	0.55	0.10	<0.001	-0.011	0.666
Sales worker (%)	0.47	0.42	0.05	0.002	0.45	0.44	0.01	0.695	0.046	0.065
Cigna insured (%)	0.41	0.19	0.21	<0.001	0.38	0.19	0.19	<0.001	0.025	0.249

Notes: The table compares the differences in demographic and job characteristics of participants and non-participants in primary data collection (Personal Health Assessment survey and/or clinical biometrics) within the 20 treatment worksites and 20 primary control worksites. The first set of columns reports differences between the participants and non-participants within the treatment worksites. The second set of columns reports differences between the participants and non-participants within primary control worksites. The last set of columns reports the difference in the differences between treatment and primary control (ΔT–ΔC). Means were estimated by linear regressions.

eTable 17. Analysis of Potential Selection: Participants vs. Non-Participants (Observational Design)

A. Self-Reported Health[a]

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Screenings and Exams			
Annual exam (%)	-0.28 (2.95) [0.92] {1.00}	7.33 (4.80) [0.14] {0.81}	-1.61 (3.62) [0.66] {1.00}
Flu shot (%)	-2.17 (3.17) [0.49] {0.99}	3.25 (5.70) [0.58] {1.00}	-3.07 (3.73) [0.41] {1.00}
Percent of other recommended tests received	4.15 (1.76) [0.02] {0.43}	5.79 (3.72) [0.14] {0.81}	4.07 (2.04) [0.05] {0.71}
Mental Health and Well-being			
PHQ-2 score of 3 or above (%)	-1.74 (1.54) [0.26] {0.97}	-4.21 (2.95) [0.17] {0.84}	-1.24 (1.57) [0.43] {1.00}
SF-8 score – physical summary score	-0.07 (0.41) [0.87] {1.00}	0.33 (0.68) [0.63] {1.00}	-0.19 (0.44) [0.66] {1.00}
SF-8 score – mental summary score	-0.44 (0.52) [0.40] {0.99}	-0.28 (1.07) [0.80] {1.00}	-0.44 (0.57) [0.43] {1.00}
Unmanaged stress (%)	-3.18 (2.62) [0.23] {0.97}	-0.03 (4.32) [0.99] {1.00}	-3.47 (3.15) [0.27] {0.99}
Stress at work (%)	3.33 (2.47) [0.18] {0.95}	8.23 (5.02) [0.12] {0.78}	2.53 (2.93) [0.39] {1.00}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Sleep			
Good quality, adequate amount of sleep (%)	-2.52 (2.19) [0.25] {0.97}	-2.91 (4.87) [0.56] {1.00}	-2.66 (2.49) [0.29] {0.99}
Physical Activity			
Regular exercise (%)	11.54 (2.14) [0.00] {0.00}	14.96 (3.61) [0.00] {0.05}	10.64 (2.75) [0.00] {0.03}
Three or more days of moderate exercise per week (%)	6.58 (2.48) [0.01] {0.30}	15.23 (3.68) [0.00] {0.05}	5.27 (2.98) [0.08] {0.84}
Number of days per week intentionally increase activity	0.14 (0.12) [0.24] {0.97}	0.39 (0.15) [0.02] {0.39}	0.11 (0.14) [0.44] {1.00}
Number of hours sitting per day	0.00 (0.11) [0.97] {1.00}	-0.15 (0.16) [0.36] {0.98}	0.03 (0.13) [0.83] {1.00}
Nutrition			
Number of meals eaten out	-0.07 (0.08) [0.40] {0.99}	-0.05 (0.09) [0.56] {1.00}	-0.07 (0.10) [0.47] {1.00}
Number of naturally or artificially sweetened drinks per day	0.02 (0.10) [0.84] {1.00}	-0.44 (0.17) [0.02] {0.39}	0.13 (0.14) [0.33] {1.00}
Read the Nutrition Facts panel (%)	6.48 (2.68) [0.02] {0.40}	10.02 (3.31) [0.01] {0.24}	5.57 (3.42) [0.10] {0.91}
Consume at least 2 cups of fruit and 2.5 cups of vegetables per day (%)	4.77 (2.15) [0.03] {0.49}	7.22 (4.32) [0.11] {0.78}	4.22 (2.76) [0.13] {0.92}
Choose whole grain foods and reduced fat foods more often than the regular variety (%)	3.23 (2.59) [0.21] {0.96}	12.67 (4.69) [0.01] {0.33}	1.55 (2.78) [0.58] {1.00}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Weight Management			
Considering losing weight in the next 6 months (%)	10.23 (2.94) [0.00] {0.07}	2.95 (3.25) [0.38] {0.98}	12.06 (3.72) [0.00] {0.11}
Actively managing weight (%)	15.22 (3.29) [0.00] {0.00}	8.97 (3.56) [0.02] {0.39}	17.24 (4.14) [0.00] {0.01}
Tobacco Use			
Smoker (%)	-8.12 (3.04) [0.01] {0.30}	-6.74 (3.24) [0.05] {0.57}	-8.78 (3.84) [0.02] {0.53}
Alcohol Use			
Number of drinks per week	-0.31 (0.33) [0.35] {0.99}	1.24 (0.47) [0.02] {0.34}	-0.73 (0.36) [0.04] [0.68]
Medical Utilization			
Number doctor visits in last 12 months	0.00 (0.06) [0.96] {1.00}	-0.03 (0.10) [0.80] {1.00}	0.00 (0.07) [0.98] {1.00}
Any doctor visit in last 12 months (%)	-1.08 (2.52) [0.67] {1.00}	-1.41 (3.80) [0.71] {1.00}	-0.75 (2.98) [0.80] {1.00}
Any ER visit in last 12 months (%)	-5.37 (2.33) [0.02] {0.44}	-8.60 (2.26) [0.00] {0.07}	-4.47 (2.87) [0.12] {0.92}
Ever hospital patient in the last 12 months (%)	-4.70 (2.11) [0.03] {0.49}	-8.28 (2.40) [0.00] {0.13}	-3.69 (2.52) [0.14] {0.93}
Days spent in hospital	-0.16 (0.07) [0.04] {0.54}	-0.32 (0.12) [0.01] {0.31}	-0.11 (0.10) [0.27] {0.99}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Number different prescriptions last 12 months	-0.07 (0.09) [0.40] {0.99}	-0.08 (0.14) [0.56] {1.00}	-0.08 (0.10) [0.43] {1.00}
Any prescriptions in last 12 months (%)	-2.12 (2.46) [0.39] {0.99}	-1.52 (5.21) [0.77] {1.00}	-2.25 (2.68) [0.40] {1.00}

B. Clinical Measures of Health[b]

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Continuous Measures			
Total Cholesterol (mg/dl)	2.49 (4.19) [0.55] {0.99}	-2.68 (3.64) [0.47] {0.93}	3.35 (5.34) [0.53] {0.99}
High Density Lipoprotein (HDL) Cholesterol (mg/dl)	0.00 (1.15) [1.00] {1.00}	1.79 (1.26) [0.17] {0.72}	-0.43 (1.32) [0.75] {1.00}
Glucose (mg/dl)	0.98 (2.88) [0.73] {1.00}	-4.75 (2.97) [0.13] {0.66}	1.79 (3.48) [0.61] {1.00}
Blood Pressure, Systolic (mm Hg)	0.49 (1.00) [0.62] {1.00}	0.74 (1.02) [0.48] {0.93}	0.30 (1.22) [0.80] {1.00}
Blood Pressure, Diastolic (mm Hg)	0.66 (0.69) [0.34] {0.94}	0.80 (0.74) [0.30] {0.81}	0.62 (0.82) [0.45] {0.98}
Body Mass Index (BMI)	0.06 (0.36) [0.86] {1.00}	-0.20 (0.56) [0.72] {0.97}	0.12 (0.44) [0.78] {1.00}
Binary Measures (%)			
High total cholesterol (total	0.20 (4.27)	-0.29 (3.50)	0.09 (5.13)

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
cholesterol ≥200 mg/dl)	[0.96] {1.00}	[0.93] {0.97}	[0.99] {1.00}
Low HDL cholesterol (HDL cholesterol <40 mg/dl)	-1.43 (2.64) [0.59] {1.00}	-1.56 (3.94) [0.70] {0.97}	-1.41 (3.03) [0.64] {1.00}
Hypertension (systolic BP ≥140 or diastolic BP ≥90 mm Hg)	3.95 (2.86) [0.17] {0.77}	4.76 (3.33) [0.17] {0.72}	3.43 (3.25) [0.29] {0.92}
Obesity (BMI ≥30)	-0.36 (2.36) [0.88] {1.00}	-5.64 (4.93) [0.27] {0.81}	0.71 (2.69) [0.79] {1.00}

C. Medical Spending and Utilization[c]

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Medical Spending ($)			
Total Spending	-644.93 (548.69) [0.24] {0.87}	-855.27 (767.77) [0.28] {0.88}	-670.13 (655.05) [0.31] {0.95}
Out-of-pocket Spending	-40.15 (55.99) [0.47] {0.98}	-72.72 (134.41) [0.59] {0.99}	-12.49 (83.13) [0.88] {1.00}
Spending By Site of Care:			
Office	-170.53 (360.81) [0.64] {0.99}	153.12 (439.83) [0.73] {0.99}	-349.59 (392.34) [0.37] {0.96}
Inpatient Hospital	-445.75 (240.39) [0.07] {0.59}	-730.66 (307.69) [0.03] {0.52}	-368.63 (369.09) [0.32] {0.95}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Emergency Room	35.87 (115.53) [0.76] {0.99}	-248.72 (142.16) [0.10] {0.71}	123.60 (144.38) [0.39] {0.97}
Urgent Care	-5.17 (4.20) [0.22] {0.87}	3.30 (4.35) [0.46] {0.97}	-9.03 (5.95) [0.13] {0.78}
Other	-59.36 (19.91) [0.00] {0.17}	-32.30 (61.06) [0.60] {0.99}	-66.48 (48.78) [0.17] {0.86}
Medical Utilization			
Utilization By Site of Care:			
Any Physician Visit (%)	3.06 (2.09) [0.14] {0.78}	8.24 (2.91) [0.01] {0.37}	0.48 (2.85) [0.87] {1.00}
Number of Physician Visits	0.18 (0.19) [0.33] {0.94}	0.20 (0.33) [0.55] {0.99}	0.17 (0.22) [0.44] {0.97}
Any Hospitalization (%)	-1.77 (1.13) [0.12] {0.74}	-2.92 (1.56) [0.08] {0.69}	-1.53 (1.36) [0.26] {0.94}
Number of Hospitalizations	-0.02 (0.01) [0.03] {0.42}	-0.03 (0.01) [0.04] {0.57}	-0.02 (0.01) [0.07] {0.64}
Any ER Visit (%)	-0.32 (2.36) [0.89] {0.99}	2.03 (3.98) [0.62] {0.99}	-0.60 (2.47) [0.81] {1.00}
Number of ER Visits	0.01 (0.04) [0.82] {0.99}	-0.07 (0.06) [0.27] {0.88}	0.04 (0.05) [0.46] {0.97}
Any Urgent Care Visit (%)	-1.27 (1.88) [0.50] {0.98}	2.12 (1.80) [0.25] {0.87}	-3.01 (2.61) [0.25] {0.94}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Number of Urgent Care Visits	-0.01 (0.03) [0.69] {0.99}	0.02 (0.03) [0.56] {0.99}	-0.03 (0.03) [0.39] {0.97}
Any Preventive Care Visit (%)	2.45 (3.75) [0.52] {0.98}	10.62 (3.90) [0.01] {0.40}	-1.00 (5.23) [0.85] {1.00}
Number of Preventive Care Visits	0.04 (0.04) [0.32] {0.93}	0.12 (0.04) [0.01] {0.30}	0.01 (0.06) [0.85] {1.00}

D. Prescription Drug Spending and Utilization[d]

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Prescription Drug Spending ($)			
Total Spending	326.53 (304.55) [0.29] {0.95}	52.12 (510.64) [0.92] {1.00}	282.50 (336.72) [0.40] {0.99}
Out-of-pocket Spending	12.77 (8.16) [0.12] {0.75}	22.46 (15.24) [0.16] {0.80}	11.09 (9.71) [0.25] {0.93}
Prescription Drug Utilization			
Any Medications (%)	2.67 (2.07) [0.20] {0.90}	2.63 (4.78) [0.59] {0.99}	3.29 (2.71) [0.22] {0.93}
Number of Distinct Medications	0.41 (0.25) [0.11] {0.74}	0.53 (0.40) [0.19] {0.85}	0.40 (0.26) [0.12] {0.80}
Number of Medication Months	0.61 (1.05) [0.57] {1.00}	0.77 (1.90) [0.69] {0.99}	0.95 (1.17) [0.42] {0.99}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
By Clinical Category:			
Any Asthma Medications (%)	(3.42) [2.05] {0.10} {0.72}	(4.59) [2.39] {0.07} {0.61}	3.22 (2.22) [0.15] {0.85}
Number of Asthma Medication Months	0.05 (0.09) [0.62] {1.00}	0.23 (0.11) [0.05] {0.54}	-0.02 (0.11) [0.86] {1.00}
Any Cardiovascular Medications (%)	1.38 (1.73) [0.43] {0.99}	3.96 (2.98) [0.20] {0.86}	0.63 (2.22) [0.78] {1.00}
Number of Cardiovascular Medication Months	0.00 (0.29) [1.00] {1.00}	0.33 (0.51) [0.52] {0.99}	-0.01 (0.37) [0.98] {1.00}
Any Diabetes Medications (%)	-0.38 (1.23) [0.76] {1.00}	-1.15 (2.10) [0.59] {0.99}	0.89 (1.40) [0.53] {1.00}
Number of Diabetes Medication Months	-0.06 (0.17) [0.71] {1.00}	-0.14 (0.31) [0.65] {0.99}	0.09 (0.24) [0.71] {1.00}
Any Hyperlipidemia Medications (%)	-0.05 (1.73) [0.98] {1.00}	1.33 (3.09) [0.67] {0.99}	-0.43 (1.71) [0.80] {1.00}
Number of Hyperlipidemia Medication Months	-0.06 (0.17) [0.71] {1.00}	-0.03 (0.31) [0.91] {1.00}	-0.06 (0.15) [0.69] {1.00}
Any Mental Health Medications (%)	1.70 (2.32) [0.46] {0.99}	1.52 (3.28) [0.65] {0.99}	1.89 (2.71) [0.48] {1.00}
Number of Mental Health Medication Months	0.12 (0.25) [0.65] {1.00}	-0.02 (0.42) [0.96] {1.00}	0.20 (0.28) [0.47] {1.00}
Any Pain Medications (%)	2.80 (1.46) [0.06] {0.60}	2.52 (3.33) [0.46] {0.98}	3.82 (2.16) [0.08] {0.68}

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Number of Pain Medication Months	0.00 (0.10) [1.00] {1.00}	0.01 (0.30) [0.98] {1.00}	0.04 (0.12) [0.75] {1.00}
Any Antibiotics Medications (%)	0.18 (1.54) [0.91] {1.00}	1.57 (2.40) [0.52] {0.99}	-0.28 (2.07) [0.89] {1.00}
Number of Antibiotics Medication Months	0.01 (0.05) [0.80] {1.00}	-0.11 (0.14) [0.44] {0.98}	0.05 (0.08) [0.57] {1.00}
Any Other Medications (%)	5.23 (1.98) [0.01] {0.23}	3.79 (4.45) [0.41] {0.97}	5.39 (2.46) [0.03] {0.45}
Number of Other Medication Months	0.56 (0.39) [0.15] {0.81}	0.51 (0.54) [0.36] {0.97}	0.66 (0.54) [0.22] {0.93}

E. Employment Outcomes[e]

	Observational Estimate Comparing Program Participants to Control Group (1)	Observational Estimate Comparing Program Participants to Non-Participants in Treatment Group (2)	RCT Estimate of Local Average Treatment Effect—Reproduced from Tables S6-S10 (3)
Absenteeism (% of scheduled hours missed)	-0.14 (0.10) [0.13] {0.31}	0.03 (0.07) [0.63] {0.61}	-0.25 (0.14) [0.08] {0.20}
Performance Review (% with a score better than 3 out of 5)	2.91 (4.18) [0.49] {0.53}	8.00 (2.66) [0.01] {0.01}	-0.78 (6.74) [0.91] {0.92}
Tenure (days employed during the treatment period)	91.00 (5.11) [0.00] {0.00}	175.16 (9.35) [0.00] {0.00}	-15.81 (19.05) [0.41] {0.45}

Notes: These tables show estimates of the treatment effects that would have been associated with the wellness program had this evaluation been done using an observational design rather than a randomized controlled design. Specifically, column (1) compared outcomes of participants in the treatment group to those of non-participants in the primary and secondary control groups and column using an observational design. Non-participants in this case were defined as individuals in the control worksites. Therefore, individuals in treatment worksites who elected not to participate where omitted from this analysis, as they could still have been exposed to the wellness program at their worksites through, for example, posters in the common areas or healthier food available in break rooms. Column (2) compared outcomes of participants to those of non-participants in the treatment group using an observational design. These may be compared to our main analyses to obtain a sense of the role of selection bias. To the extent that these observational results differ from results derived from the randomized trial design, they highlight the importance of using a randomized approach to evaluation for such programs when participation is voluntary.

Standard errors are listed in parentheses with p-values in brackets and family-wise p-values in curly braces. The sample sizes provided are for the intent-to-treat sample only. For the domains of self-reported health and clinical measures of health, the sample sizes differ by outcome based on the number of respondents or data participants for each outcome. The sample sizes are identical for the domains of medical and pharmaceutical claims outcomes, as they are derived from the same subset of individuals with Cigna coverage. The sample sizes for the domain of employment outcomes differ based on the availability of absenteeism, performance review, and tenure data at the individual level.

[a] Sample sizes in the observational analysis comparing program participants to control group (1) ranged between 1,465 and 1,717. Sample sizes in the observational analysis comparing program participants and non-participants in treatment group (2) ranged between 864 and 1,013.

[b] Sample sizes in the observational analysis comparing program participants to control group (1) ranged between 1,752 and 1,795. Sample sizes in the observational analysis comparing program participants and non-participants in treatment group (2) ranged between 1,036 and 1,065.

[c, d] Sample size in the observational analysis comparing program participants to control group (1) was 7,217. Sample size in the observational analysis comparing program participants and non-participants in treatment group (2) was 1,005.

[e] Sample sizes in the observational analysis comparing program participants to control group (1) ranged between 122,446 and 30,356. Sample sizes in the observational analysis comparing program participants and non-participants in treatment group (2) ranged between 2,975 and 4,037.

